```
            YE  TLEK+ +LRAQVA A++PA+AM AS+ LTRQ+S I AVAE+YPDLKAN +++KLQ
Sbjct: 61  KYEQATLEKVTQLRAQVASASSPADAMKASDALTRQISGIFAVAESYPDLKANENYLKLQ 120

Query: 121 EELTNTENKISYSRQLYNTTTSNYNVKLETFPSNIVGKLFGFKPSQFLETPEEEKEVPKV 180
            EELTNTENKISYSRQLYN+   NYNVKL+ FPSN++  +F F+P+ FL TPEEEK VPKV
Sbjct: 121 EELTNTENKISYSRQLYNSVAGNYNVKLQAFPSNVIAGMFAFRPADFLSTPEEEKAVPKV 180

Query: 181 SF 182
             F
Sbjct: 181 DF 182
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4857> which encodes the amino acid sequence <SEQ ID 4858>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC44350 GB:U66186 LemA [Listeria monocytogenes]
 Identities = 91/181 (50%), Positives = 121/181 (66%), Gaps = 2/181 (1%)

Query: 5   LIILVVLGVLALWLMISYNSLVKSRMHTKEAWSQIDVQLKRRNDLIPNLIETVKGYASYE 64
            +I + V+ +L L     YNSLVK R    E W+QIDVQLKRR DLIPNL+ETVKGYA +E
Sbjct: 5   IIAIAVVVILVLIYFGLYNSLVKYRNRVDETWAQIDVQLKRRFDLIPNLVETVKGYAKHE 64

Query: 65  QKTFEKITDLRARVAN--ASTPQETMAASNELSKQVTSLFAVAENYPDLKANENFLKLQE 122
            ++T  ++ + R ++    A   Q  + A N LS  + S+FA+ E YPDLKAN +F++LQ
Sbjct: 65  KETLTQVIEARNKMMEVPADNRQGQIEADNMLSGALKSIFALGEAYPDLKANTSFIELQH 124

Query: 123 ELTNTENKISYSRQLYNSTTSNYNLQLESFPSNIAGKLFGFKPSEFLQTPEAEKEVPKVEF 183
            ELT TENK++YSRQLYN+T   YN +++S P+NI  KL  F   + L  PE E+  PKVEF
Sbjct: 125 ELTTTENKVAYSRQLYNTTVMTYNTKVQSVPTNIVAKLHNFTERDMLSIPEVERVAPKVEF 185
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 135/181 (74%), Positives = 165/181 (90%)

Query: 4   MILIAIIALFVIWLIVAYNSLVRSRMHTKESWSQIDVQLKRRNDLIPNLIETVKGYAAYE 63
            +I++ ++ +  +WL+++YNSLV+SRMHTKE+WSQIDVQLKRRNDLIPNLIETVKGYA+YE
Sbjct: 5   LIILVVLGVLALWLMISYNSLVKSRMHTKEAWSQIDVQLKRRNDLIPNLIETVKGYASYE 64

Query: 64  GKTLEKIAELRAQVAKANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSFVKLQEEL 123
             KT EKI +LRA+VA A+TP E M ASNEL++Q++S+ AVAENYPDLKAN +F+KLQEEL
Sbjct: 65  QKTFEKITDLRARVANASTPQETMAASNELSKQVTSLFAVAENYPDLKANENFLKLQEEL 124

Query: 124 TNTENKISYSRQLYNTTTSNYNVKLETFPSNIVGKLFGFKPSQFLETPEEEKEVPKVSFDF 184
            TNTENKISYSRQLYN+TTSNYN++LE+FPSNI GKLFGFKPS+FL+TPE EKEVPKV F+F
Sbjct: 125 TNTENKISYSRQLYNSTTSNYNLQLESFPSNIAGKLFGFKPSEFLQTPEAEKEVPKVEFNF 185
```

A related GBS gene <SEQ ID 8849> and protein <SEQ ID 8850> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1    Crend: 0
McG: Discrim Score:     14.63
GvH: Signal Score (-7.5): -3.19
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
ALOM program    count: 1 value: -15.44 threshold:  0.0
   INTEGRAL     Likelihood =-15.44    Transmembrane     4 -  20 (   1 -  27)
```

```
  PERIPHERAL  Likelihood =  8.86     146
 modified ALOM score:   3.59

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
51.4/68.9% over 183aa
                       Listeria monocytogenes
  EGAD|149857| LemA protein Insert characterized
  GP|1519287|gb|AAC44350.1||U66186 LemA Insert characterized

ORF01545(301 - 846 of 1152)
EGAD|149857|159923(2 - 185 of 185) LemA protein {Listeria monocytogenes}
GP|1519287|gb|AAC44350.1||U66186 LemA {Listeria monocytogenes}
%Match = 23.8
%Identity = 51.4  %Similarity = 68.9
Matches = 94  Mismatches = 56  Conservative Sub.s = 32

42        72        102       132       162       192       222       252
CFK*TSSLSVIAVRLIFSFHSTRSLK*VSNCFFCLSVSVIPCSIRT**NAWGVIVNLNFYIV**LYFITNTNNGNNRTFL

282       312       342       372       402       432       462       492
I*RKLL*WKKCKGATTMGTMILIAIIALFVIWLIVAYNSLVRSRMHTKESWSQIDVQLKRRNDLIPNLIETVKGYAAYEG
                :| :| ||:: ::|:     |||||: |    |:|:||||||||| ||||||:||||||| :|
               MIGWIIAIAVVVILVLIYFGLYNSLVKYRNRVDETWAQIDVQLKRRFDLIPNLVETVKGYAKHEK
                       10        20        30        40        50        60

522       546       576       606       636       666       696       726
KTLEKIAELRAQVAK--ANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSFVKLQEELTNTENKISYSRQLYNTTTS
:|| :: | | :: :  |:     : | | |:  | ||:|: | ||||||| ||::|| ||| ||||::|||||||||
ETLTQVIEARNKMMEVPADNRQGQIEADNMLSGALKSIFALGEAYPDLKANTSFIELQHELTTTENKVAYSRQLYNTTVM
             80        90        100       110       120       130       140

756       786       816       846       876       906       936       966
NYNVKLETFPSNIVGKLFGFKPSQFLETPEEEKEVPKVSFDF*LRRERGFCCINKLQVIREKQLSC*LSSSVF*QLLEQL
 || |::: |:||| ||  |     |  || |:  ||| |
TYNTKVQSVPTNIVAKLHNFTERDMLSIPEVERVAPKVEF
            160       170       180
```

SEQ ID 4856 (GBS42) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 2; MW 21.8kDa) and in Figure 168 (lane 5-7; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 8; MW 46kDa). Purified Thio-GBS42-His is shown in Figure 244, lane 11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1572

A DNA sequence (GBSx1666) was identified in *S.agalactiae* <SEQ ID 4859> which encodes the amino acid sequence <SEQ ID 4860>. This protein is predicted to be glucose inhibited division protein b (gidB). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                bacterial cytoplasm --- Certainty=0.2430(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10079> which encodes amino acid sequence <SEQ ID 10080> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB16137 GB:Z99124 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 130/239 (54%), Positives = 170/239 (70%), Gaps = 4/239 (1%)

Query: 5   MTPQAFYQVLIEHGITLTDKQKKQFETYFRLLVEWNEKINLTAITDKEEVYLKHFYDSIA 64
            M  + F   L E GI+L+ +Q +QFE Y+ +LVEWNEKINLT+IT+K+EVYLKHFYDSI
Sbjct: 1   MNIEEFTSGLAEKGISLSPRQLEQFELYYDMLVEWNEKINLTSITEKKEVYLKHFYDSIT 60

Query: 65  PILQGYID-NSPLSILDIGAGAGFPSIPMKILYPEIDITIIDSLNKRINFLNILANELEL 123
                 Y+D N   +I D+GAGAGFPS+P+KI +P + +TI+DSLNKRI FL  L+  L+L
Sbjct: 61  AAF--YVDFNQVNTICDVGAGAGFPSLPIKICFPHLHVTIVDSLNKRITFLEKLSEALQL 118

Query: 124 SGVHFFHGRAEDFGQDRVFRAKFDIVTARAVAKMQVLAELTIPFLKVNGRLIALKAAAAE 183
                F H RAE FGQ +  R  +DIVTARAVA++ VL+EL +P +K NG  +ALKAA+AE
Sbjct: 119 ENTTFCHDRAETFGQRKDVRESYDIVTARAVARLSVLSELCLPLVKKNGLFVALKAASAE 178

Query: 184 EELISAEKALKTLFSQVTVNKNYKLP-NGDDRNITIVSKKKETPNKYPRKAGTPNKKPL 241
            EEL + +KA+ TL  ++    ++KLP    DRNI ++ K K TP KYPRK GTPNK P+
Sbjct: 179 EELNAGKKAITTLGGELENIHSFKLPIEESDRNIMVIRKIKNTPKKYPRKPGTPNKSPI 237
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4861> which encodes the amino acid sequence <SEQ ID 4862>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4862(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 170/237 (71%), Positives = 202/237 (84%)

Query: 5   MTPQAFYQVLIEHGITLTDKQKKQFETYFRLLVEWNEKINLTAITDKEEVYLKHFYDSIA 64
            MTPQ FY+ L E G +L+ KQK+QF+TYF+ LVEWN KINLTAIT++ EVYLKHFYDSIA
Sbjct: 1   MTPQDFYRTLEEDGFSLSSKQKEQFDTYFKSLVEWNTKINLTAITEENEVYLKHFYDSIA 60

Query: 65  PILQGYIDNSPLSILDIGAGAGFPSIPMKILYPEIDITIIDSLNKRINFLNILANELELS 124
            PILQG++ N P+ +LDIGAGAGFPS+PMKIL+P +++TIIDSLNKRI+FL +LA EL L
Sbjct: 61  PILQGFLANEPIKLLDIGAGAGFPSLPMKILFPNLEVTIIDSLNKRISFLTLLAQELGLE 120

Query: 125 GVHFFHGRAEDFGQDRVFRAKFDIVTARAVAKMQVLAELTIPFLKVNGRLIALKAAAAEE 184
             VHFFHGRAEDFGQD+ FR +FD+VTARAVA+MQVL+ELTIPFLK+ G+LIALKA AA++
Sbjct: 121 NVHFFHGRAEDFGQDKAFRGQFDVVTARAVARMQVLSELTIPFLKIGGKLIALKAQAADQ 180

Query: 185 ELISAEKALKTLFSQVTVNKNYKLPNGDDRNITIVSKKKETPNKYPRKAGTPNKKPL 241
            EL  A+ AL  LF +V  N +Y+LPNGD R ITIV KKKETPNKYPRKAG PNKKPL
Sbjct: 181 ELEEAKNALCLLFGKVIKNHSYQLPNGDSRFITIVEKKKETPNKYPRKAGLPNKKPL 237
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1573

A DNA sequence (GBSx1667) was identified in *S.agalactiae* <SEQ ID 4863> which encodes the amino acid sequence <SEQ ID 4864>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1574

A DNA sequence (GBSx1668) was identified in *S.agalactiae* <SEQ ID 4865> which encodes the amino acid sequence <SEQ ID 4866>. This protein is predicted to be v-type sodium ATP synthase subunit j. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.14   Transmembrane  371 - 387 ( 362 - 391)
   INTEGRAL    Likelihood = -7.48   Transmembrane  200 - 216 ( 190 - 217)
   INTEGRAL    Likelihood = -4.94   Transmembrane  425 - 441 ( 423 - 446)
   INTEGRAL    Likelihood = -4.67   Transmembrane  327 - 343 ( 325 - 349)
   INTEGRAL    Likelihood = -3.77   Transmembrane   81 -  97 (  81 -  98)
   INTEGRAL    Likelihood = -2.66   Transmembrane  140 - 156 ( 139 - 157)
   INTEGRAL    Likelihood = -1.33   Transmembrane   55 -  71 (  53 -  71)
   INTEGRAL    Likelihood = -0.27   Transmembrane  247 - 263 ( 247 - 263)
   INTEGRAL    Likelihood = -0.11   Transmembrane  165 - 181 ( 165 - 181)

----- Final Results -----
              bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10081> which encodes amino acid sequence <SEQ ID 10082> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]
 Identities = 170/461 (36%), Positives = 262/461 (55%), Gaps = 28/461 (6%)

Query: 12  KTMSVARKLSISFIAVILLGSILLSLPIFQYANAPKTHYIDHLFTTVSMVCVTGLSVFPI 71
            K +S  + ++  F  +IL G  LL+LP F   +   TH+ID LFT  S VCVTGL+
Sbjct: 10  KRLSPVQLIAAGFFILILFGGSLLTLPFFS-RSGESTHFIDALFTATSAVCVTGLTTLNT 68

Query: 72  SKVYNGWGQIVAILLMQTGGLGLVTLMSLSYYTLRRKMSLNDQTLLQSAITYNSSTDLKK 131
            ++ +N  GQ + + L++ GGLG + +  L +   ++K+S + + +L+ A+     + + K
Sbjct: 69  AEHWNSAGQFLIMTLIEIGGLGFMMIPILFFAIAKKKISFSMRIVLKEALNLEEMSGVIK 128

Query: 132 YLYMIFKVTLTLEVLAASILAIDFIPRFGLGHGIFNSIFLAVSAFCNAGFDNLEATSLAQ 191
             +  I K  + ++V+ A  L++ FIP FG   GI+ SIF AVS+FCNAGFD L  + LA
Sbjct: 129 LMIYILKFAVVIQVIGAVALSVVFIPEFGWAKGIWFSIFHAVSSFCNAGFDLLGDSLLAD 188
```

```
Query: 192 FKLNPLVNIIVCFLIISGGLGFAVWKDLIEATIQTSHKGPKLIKTFPKRLSNHSKLVLKT 251
            + N  + ++V  LII+GGLGF VW+D++     + H+         K+++ HSK+ L
Sbjct: 189 -QTNVYLIMVVSALIIAGGLGFIVWRDIL-----SYHR--------VKKITLHSKVALSV 234

Query: 252 TTIILLTGTLLSWLLEFGNFRTIANLSLPKQLMVSFFQTVTMRTAGFSTIDYTQTDFATN 311
           T ++L+ G +L +L+   N  T+   +  ++L  +FF +VT RTAG+ +IDY Q   A
Sbjct: 235 TALLLIGGFIL-FLITERNGLTLVKGTFTERLANTFFMSVTPRTAGYYSIDYLQMSHAGL 293

Query: 312 LVYIIQMLIGGAPGGTAGGFKVTVIAILLLLFKAELSGQSQVTFHYRTIPSSIIKQTLSI 371
           ++ +  M IGG  G TAGG K T + ILL+   A   G+++     RTI  + +   L
Sbjct: 294 ILTMFLMYIGGTSGSTAGGLKTTTLGILLIQMHAMFKGKTRAEAFGRTIRQAAV---LRA 350

Query: 372 LTFFFII--LISGYLLLLELNPHIDPFS----LFFEASSALATVGVTMNTTNQLTLGGRI 425
           LT FF+   L    +++L +   I   S    + FE  SA  TVG+TM  T  LTL G++
Sbjct: 351 LTLFFVTLSLCVVAIMVLSVTETIPKTSGIEYIAFEVFSAFGTVGLTMGLTPDLTLIGKL 410

Query: 426 VIMFLMFIGRVGPITVLLSILQK---KEKEIHYAETEIILG 463
           VI+ LM+IGRVG +TV+LS+L K    E    Y E  I+LG
Sbjct: 411 VIISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG 451
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4867> which encodes the amino acid sequence <SEQ ID 4868>. Analysis of this protein sequence reveals the following:

```
     Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.12   Transmembrane  371 - 387 ( 364 - 396)
   INTEGRAL    Likelihood = -7.32   Transmembrane   20 -  36 (  18 -  42)
   INTEGRAL    Likelihood = -6.53   Transmembrane  425 - 441 ( 417 - 446)
   INTEGRAL    Likelihood = -6.16   Transmembrane   89 - 105 (  81 - 106)
   INTEGRAL    Likelihood = -5.79   Transmembrane  200 - 216 ( 196 - 223)
   INTEGRAL    Likelihood = -3.35   Transmembrane  140 - 156 ( 139 - 157)
   INTEGRAL    Likelihood = -3.03   Transmembrane   55 -  71 (  53 -  74)
   INTEGRAL    Likelihood = -3.03   Transmembrane  247 - 263 ( 246 - 264)
   INTEGRAL    Likelihood = -1.12   Transmembrane  393 - 409 ( 393 - 409)
   INTEGRAL    Likelihood = -0.11   Transmembrane  165 - 181 ( 165 - 181)

----- Final Results -----
           bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]
 Identities = 168/466 (36%), Positives = 260/466 (55%), Gaps = 26/466 (5%)

Query: 6   MKRSFIKSLSVTQRLTFSFAIVILIGTLLLSMPFTHYQNGPNTVYLDHFFNVVSMVCVTG 65
           MK+   K LS  Q +   F I+IL G  LL++PF   ++G +T ++D  F   S VCVTG
Sbjct: 4   MKKRVRKRLSPVQLIAAGFFILILFGGSLLTLPFFS-RSGESTHFIDALFTATSAVCVTG 62

Query: 66  LSVVPVAEVYNGIGQTIAMALMQIGCLGLVTLIAVSTFAL-KRKMRLSDQTLLQSALNRG 124
           L+ +  AE +N  GQ + M L++IG LG + +I +  FA+ K+K+  S + +L+ ALN
Sbjct: 63  LTTLNTAEHWNSAGQFLIMTLIEIGGLGFM-MIPILFFAIAKKKISFSMRIVLKEALNLE 121

Query: 125 DSKDLKHYLFFAYKVTFSLEAFAAIVIMIDFIPRFGWKNGIFNSIFLAVSAFCNAGFDNL 184
           +   +   + +  K    ++   A+ + + FIP FGW  GI+ SIF AVS+FCNAGFD L
Sbjct: 122 EMSGVIKLMIYILKFAVVIQVIGAVALSVVFIPEFGWAKGIWFSIFHAVSSFCNAGFDLL 181

Query: 185 GSSSLKDFMLNPTLNVIITFLIISGGLGFAVWVDLGVAFKKYFFERPHCYGATFRKLSNQ 244
           G S L D   N  L ++++ LII+GGLGF VW D+ +++ +             +K++
Sbjct: 182 GDSLLAD-QTNVYLIMVVSALIIAGGLGFIVWRDI-LSYHR------------VKKITLH 227

Query: 245 SRLVLQTTAVILFLGTFLTWFLEKDNSKTIANFSLHQQLMVSFFQTVTMRTAGFATISYN 304
           S++ L  TA++L +G F+ + + + N  T+   +  ++L  +FF +VT RTAG+ +I Y
Sbjct: 228 SKVALSVTALLL-IGGFILFLITERNGLTLVKGTFTERLANTFFMSVTPRTAGYYSIDYL 286

Query: 305 DTLAPTNILYMIQMVIGGAPGGTAGGIKVTTAAITFLLFKAELSGQSEVTFRNRIIANKT 364
```

```
                  IL M  M IGG  G TAGG+K TT  I  +   A   G++      R I
Sbjct: 287 QMSHAGLILTMFLMYIGGTSGSTAGGLKTTTLGILLIQMHAMFKGKTRAEAFGRTIRQAA 346

Query: 365 IKQTMTVLIFFFAVLMIGFILLLSVEPHIAPIP----LLFESISAIATVGVSMDLTPQLS 420
           + + +T L F    L +  I++LSV   I        + FE  SA  TVG++M LTP L+
Sbjct: 347 VLRALT-LFFVTLSLCVVAIMVLSVTETIPKTSGIEYIAFEVFSAFGTVGLTMGLTPDLT 405

Query: 421 TAGRLIVIVLMFVGRVGPITVLISLI---QRKEKTIQYATTDILVG 463
             G+L++I LM++GRVG +TV++SL+    R E   +Y    I++G
Sbjct: 406 LIGKLVIISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG 451
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 275/462 (59%), Positives = 351/462 (75%), Gaps = 1/462 (0%)

Query: 2   GASMKHFFDYKTMSVARKLSISFIAVILLGSILLSLPIFQYANAPKTHYIDHLFTTVSMV 61
           G +MK  F  K++SV ++L+ SF  VIL+G++LLS+P   Y N P T Y+DH F  VSMV
Sbjct: 3   GGNMKRSF-IKSLSVTQRLTFSFAIVILIGTLLLSMPFTHYQNGPNTVYLDHFFNVVSMV 61

Query: 62  CVTGLSVFPISKVYNGWGQIVAILLMQTGGLGLVTLMSLSYYTLRRKMSLNDQTLLQSAI 121
           CVTGLSV P+++VYNG GQ +A+ LMQ G LGLVTL+++S + L+RKM L+DQTLLQSA+
Sbjct: 62  CVTGLSVVPVAEVYNGIGQTIAMALMQIGCLGLVTLIAVSTFALKRKMRLSDQTLLQSAL 121

Query: 122 TYNSSTDLKKYLYMIFKVTLTLEVLAASILAIDFIPRFGLGHGIFNSIFLAVSAFCNAGF 181
               S DLK YL+  +KVT +LE  AA ++ IDFIPRFG  +GIFNSIFLAVSAFCNAGF
Sbjct: 122 NRGDSKDLKHYLFFAYKVTFSLEAFAAIVIMIDFIPRFGWKNGIFNSIFLAVSAFCNAGF 181

Query: 182 DNLEATSLAQFKLNPLVNIIVCFLIISGGLGFAVWKDLIEATIQTSHKGPKLIKTFPKRL 241
           DNL ++SL  F LNP +N+I+ FLIISGGLGFAVW DL  A  +   + P       ++L
Sbjct: 182 DNLGSSSLKDFMLNPTLNVIITFLIISGGLGFAVWVDLGVAFKKYFFERPHCYGATFRKL 241

Query: 242 SNHSKLVLKTTTIILLTGTLLSWLLEFGNFRTIANLSLPKQLMVSFFQTVTMRTAGFSTI 301
           SN S+LVL+TT +IL  GT L+W LE  N +TIAN SL +QLMVSFFQTVTMRTAGF+TI
Sbjct: 242 SNQSRLVLQTTAVILFLGTFLTWFLEKDNSKTIANFSLHQQLMVSFFQTVTMRTAGFATI 301

Query: 302 DYTQTDFATNLVYIIQMLIGGAPGGTAGGFKVTVIAILLLLFKAELSGQSQVTFHYRTIP 361
            Y  T   TN++Y+IQM+IGGAPGGTAGG KVT  AI  LLFKAELSGQS+VTF  R I
Sbjct: 302 SYNDTLAPTNILYMIQMVIGGAPGGTAGGIKVTTAAITFLLFKAELSGQSEVTFRNRIIA 361

Query: 362 SSIIKQTLSILTFFFIILISGYLLLLELNPHIDPFSLFFEASSALATVGVTMNTTNQLTL 421
           +  IKQT+++L FFF +L+ G++LLL + PHI P  L FE+ SA+ATVGV+M+ T QL+
Sbjct: 362 NKTIKQTMTVLIFFFAVLMIGFILLLSVEPHIAPIPLLFESISAIATVGVSMDLTPQLST 421

Query: 422 GGRIVIMFLMFIGRVGPITVLLSILQKKEKEIHYAETEIILG 463
            GR++++ LMF+GRVGPITVL+S++Q+KEK I YA T+I++G
Sbjct: 422 AGRLIVIVLMFVGRVGPITVLISLIQRKEKTIQYATTDILVG 463
```

A related GBS gene <SEQ ID 8851> and protein <SEQ ID 8852> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      0.86
GvH: Signal Score (-7.5): 0.64
     Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 9 value: -10.14 threshold:  0.0
   INTEGRAL    Likelihood =-10.14   Transmembrane  371 - 387 ( 362 - 391)
   INTEGRAL    Likelihood = -7.48   Transmembrane  200 - 216 ( 190 - 217)
   INTEGRAL    Likelihood = -4.94   Transmembrane  425 - 441 ( 423 - 446)
   INTEGRAL    Likelihood = -4.67   Transmembrane  327 - 343 ( 325 - 349)
   INTEGRAL    Likelihood = -3.77   Transmembrane   81 -  97 (  81 -  98)
   INTEGRAL    Likelihood = -2.66   Transmembrane  140 - 156 ( 139 - 157)
   INTEGRAL    Likelihood = -1.33   Transmembrane   55 -  71 (  53 -  71)
   INTEGRAL    Likelihood = -0.27   Transmembrane  247 - 263 ( 247 - 263)
   INTEGRAL    Likelihood = -0.11   Transmembrane  165 - 181 ( 165 - 181)
   PERIPHERAL  Likelihood =  2.49     308
 modified ALOM score:   2.53
```

```
*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02334(334 - 1689 of 1989)
EGAD|22151|22827(10 - 451 of 451) v-type sodium ATP synthase subunit j {Enterococcus hirae}
SP|P43440|NTPJ_ENTHR V-TYPE SODIUM ATP SYNTHASE SUBUNIT J (EC 3.6.1.34) (NA(+)-
TRANSLOCATING ATPASE SUBUNIT J). GP|487282|dbj|BAA04279.1||D17462 Na+ -ATPase subunit J
{Enterococcus hirae}
%Match = 18.8
%Identity = 38.5  %Similarity = 60.4
Matches = 170  Mismatches = 166  Conservative Sub.s = 97

186       216       246       276       306       336       366       396
TIFTSNCK*KL*VT*W**PKYHNR*QEKRNA**IPS*SWYSKQEAFVKLGASMKHFFDYKTMSVARKLSISFIAVILLGS
                                                           | :|   : ::  |  :||:|
                                                  MTIMKKRVRKRLSPVQLIAAGFFILILFGG
                                                          10        20        30

426       456       486       516       546       576       606       636
ILLSLPIFQYANAPKTHYIDHLFTTVSMVCVTGLSVFPISKVYNGWGQIVAILLMQTGGLGLVTLMSLSYYTLRRKMSLN
 ||:|| |  :    ||:|| |||  | ||||||: :  :: :|  || : : |:: ||||:: :  | :   ::|:|::
SLLTLPFFSRSGES-THFIDALFTATSAVCVTGLTTLNTAEHWNSAGQFLIMTLIEIGGLGFMMIPILFFAIAKKKISFS
        40         50         60         70         80         90       100

666       696       726       756       786       816       846       876
DQTLLQSAITYNSSTDLKKYLYMIFKVTLTLEVLAASILAIDFIPRFGLGHGIFNSIFLAVSAFCNAGFDNLEATSLAQF
 : :|: |:     : : | :  |:|  : ::|: |  |:: ||| ||   ||: ||| |||:||||||| |   ||
MRIVLKEALNLEEMSGVIKLMIYILKFAVVIQVIGAVALSVVFIPEFGWAKGIWFSIFHAVSSFCNAGFD-LLGDSLLAD
        120       130       140       150       160       170        180

906       936       966       996       1026      1056      1086      1116
KLNPLVNIIVCFLIISGGLGFAVWKDLIEATIQTSHKGPKLIKTFPKRLSNHSKLVLKTTTIILLTGTLLSWLLEFGNFR
: |  : ::|  |||:||||| ||:|:             |     |::: |||: |  |  :|| | :: :|:   |
QTNVYLIMVVSALIIAGGLGFIVWRDI-------------LSYHRVKKITLHSKVALSVTA-LLLIGGFILFLITERNGL
         200       210                    220       230        240       250

1146      1176      1206      1236      1266      1296      1326      1356
TIANLSLPKQLMVSFFQTVTMRTAGFSTIDYTQTDFATNLVYIIQMLIGGAPGGTAGGFKVTVIAILLLLFKAELSGQSQ
|:   :: ::|  :|| :|| ||||: :||| |   |  :: :  | |||  | ||||:| | : |||:   | : |:::
TLVKGTFTERLANTFFMSVTPRTAGYYSIDYLQMSHAGLILTMFLMYIGGTSGSTAGGLKTTTLGILLIQMHAMFKGKTR
             270       280       290       300       310       320       330

1386      1416                1461      1491      1518      1548      1578
VTFHYRTIPSSIIKQTLSILTFFFIIL----ISGYLL-LLELNPHIDPFS-LFFEASSALATVGVTMNTTNQLTLGGRIV
     |||  : :   |  ||:||: |    ::  :| : |  |       : ||  ||: |||:||  |  ||| |::|
AEAFGRTIRQAAV---LRALTLFFVTLSLCVVAIMVLSVTETIPKTSGIEYIAFEVFSAFGTVGLTMGLTPDLTLIGKLV
                350       360       370       380       390       400       410

1608      1638      1659      1689      1719      1749      1779      1809
IMFLMFIGRVGPITVLLSILQK---KEKEIHYAETEIILG*KRSFMKTKIIGVLGLGIFGQTLAQELSNFEQDVIAIDSN
|: ||:||||| :||:||:| |    |    | |  |:||
IISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG
                430       440       450
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1575

A DNA sequence (GBSx1669) was identified in *S.agalactiae* <SEQ ID 4869> which encodes the amino acid sequence <SEQ ID 4870>. This protein is predicted to be TrkA (ktrA). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC46144 GB:AF001974 putative TrkA [Thermoanaerobacter
           ethanolicus]
 Identities = 69/177 (38%), Positives = 110/177 (61%), Gaps = 2/177 (1%)

Query: 8   VLGLGIFGQTLAQELSNFEQDVIAIDSNPEN--VQAVAEVVTKAAIGDITDLAFLKHIGI 65
           V+GLG FG +LA+ L     DV+ ID + E   VQA+  +VT A   D TD   LK + +
Sbjct: 6   VIGLGSFGISLAKTLYEMGNDVLVIDEDEEEELVQAMNGLVTHAVRADATDENVLKSLRV 65

Query: 66  SDCDTVIIATGNSLESSVLAVMHCKKLGVPQVIAKARNLVYEEVLYEIGADLVISPERES 125
            + D  I+A G ++ESS++  M  K+LGV  VIAKA N ++  VLY++GAD V+ PE++
Sbjct: 66  KNFDVAIVAIGKNMESSIMVTMLVKELGVKYVIAKAHNELHARVLYKVGADRVVMPEKDM 125

Query: 126 GQNVAANLMRNKITDVFQIESDISVIEFKIPKSWVGKTVEQLNIRHKFDLNLIGIRK 182
           G  VA N+  + + D+ +   + S+ E    + W GKT++++N+R K+ LN++ ++K
Sbjct: 126 GIRVARNVFSSNLIDLIEFSKEYSIAEILPIEEWFGKTLKEINVREKYGLNVVAVKK 182
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4715> which encodes the amino acid sequence <SEQ ID 4716>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 132/221 (59%), Positives = 176/221 (78%)

Query: 1   MKTKIIGVLGLGIFGQTLAQELSNFEQDVIAIDSNPENVQAVAEVVTKAAIGDITDLAFL 60
           +K K +GVLGLGIFG+T+A+ELSNF+QDVIAID    +V+ VA++VTKAA+GDITD  FL
Sbjct: 2   LKRKTVGVLGLGIFGRTVARELSNFDQDVIAIDIRESHVKEVADLVTKAAVGDITDKEFL 61

Query: 61  KHIGISDCDTVIIATGNSLESSVLAVMHCKKLGVPQVIAKARNLVYEEVLYEIGADLVIS 120
             +GI  CDTV+IA+GN+LESSVLAVMHCKKLGVP +IAKA+N ++EEVLY IGA  VI+
Sbjct: 62  LAVGIEHCDTVVIASGNNLESSVLAVMHCKKLGVPTIIAKAKNKIFEEVLYGIGATKVIT 121

Query: 121 PERESGQNVAANLMRNKITDVFQIESDISVIEFKIPKSWVGKTVEQLNIRHKFDLNLIGI 180
           PER+SG+ VA+NL+R  I  +  +E  IS+IEF IPKSW G+++ +L++R K++LN+IG+
Sbjct: 122 PERDSGKRVASNLLRRHIESIIYLEHGISMIEFVIPKSWEGQSLSELDVRRKYELNVIGM 181

Query: 181 RKAKNKPVDTEVPINSPLEEGIILVAIANSDAFQRYDYLGY 221
           R+ + K +DT V    PLE   I+VAIAN   F+++DYLGY
Sbjct: 182 RQKEVKTLDTNVKPFEPLEPNTIIVAIANDHTFEKFDYLGY 222
```

A related GBS gene <SEQ ID 8853> and protein <SEQ ID 8854> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      5.14
GvH: Signal Score (-7.5): -0.860001
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   1.06 threshold:  0.0
   PERIPHERAL  Likelihood =  1.06     192
 modified ALOM score:  -0.71

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

The protein has homology with the following sequences in the databases:

```
38.0/61.6% over 182aa
                                                                   Thermoanaerobacter
ethanolicus
  GP|2581796| putative TrkA Insert characterized

ORF02030(322 - 864 of 1269)
GP|2581796|gb|AAC46144.1||AF001974(6  -  188  of  195)  putative  TrkA  {Thermoanaerobacter
ethanolicus}
%Match = 15.5
%Identity = 37.9  %Similarity = 61.5
Matches = 69  Mismatches = 69  Conservative Sub.s = 43

60        90       120       150       180       210       240       270
LISGYLLLLELNPHIDPFSLFFEASSALATVGVTMNTTNQLTLGGRIVIMFLMFIGRVGPITVLLSILQKKEKEIHYAET

300       330       360       390                 444       474       504
EIILG*KRSFMKTKIIGVLGLGIFGQTLAQELSNFEQDVIAIDSNPEN--VQAVAEVVTKAAIGDITDLAFLKHIGISDC
             |   |:||| || :||: |     ||: || : |   |||:  :|| |   | ||   || : : :
            MKQFVVIGLGSFGISLAKTLYEMGNDVLVIDEDEEEELVQAMNGLVTHAVRADATDENVLKSLRVKNF
                    10        20        30        40        50        60

534       564       594       624       654       684       714       744
DTVIIATGNSLESSVLAVMHCKKLGVPQVIAKARNLVXEEVLYEIGADLVISPERESGQNVAANLMRNKITDVFQIESDI
|  |:| | ::|||::  |  |:|||  ||||| | :   |||::||| |: ||:: |  || |:  : : |: :   :
DVAIVAIGKNMESSIMVTMLVKELGVKYVIAKAHNELHARVLYKVGADRVVMPEKDMGIRVARNVFSSNLIDLIEFSKEY
          80        90       100       110       120       130       140

774       804       834       864       894       924       954       984
SVIEFKIPKSWVGKTVEQLNIRHKFDLNLIGIRKAKNKPVDTEVPINSPLEEXIILVAIANSDAFQRYDYLRYFY*RK*K
|: |    : | |||::::|:| |: ||:: ::|  :: : :
SIAEILPIEEWFGKTLKEINVREKYGLNVVAVKKFNDEIIVSPGAGL
         160       170       180       190
```

SEQ ID 8854 (GBS57) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 6; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 11; MW 51.1kDa) and in Figure 183 (lane 9 & 10; MW 51kDa).

The GBS57-GST fusion product was purified (Figure 99A; see also Figure 195, lane 8) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 99B), FACS (Figure 99C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1576

A DNA sequence (GBSx1670) was identified in *S.agalactiae* <SEQ ID 4871> which encodes the amino acid sequence <SEQ ID 4872>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.62   Transmembrane   73 -  89 (  68 -  96)
   INTEGRAL    Likelihood =-11.30   Transmembrane  254 - 270 ( 248 - 274)
   INTEGRAL    Likelihood = -4.73   Transmembrane  127 - 143 ( 124 - 144)
   INTEGRAL    Likelihood = -4.19   Transmembrane   50 -  66 (  47 -  67)
   INTEGRAL    Likelihood = -3.29   Transmembrane   25 -  41 (  25 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8855> which encodes amino acid sequence <SEQ ID 8856> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 9
McG: Discrim Score:    -10.49
GvH: Signal Score (-7.5): -1.14
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 5 value: -11.62 threshold:  0.0
   INTEGRAL    Likelihood =-11.62   Transmembrane   73 -  89 (  68 -  96)
   INTEGRAL    Likelihood =-11.30   Transmembrane  254 - 270 ( 248 - 274)
   INTEGRAL    Likelihood = -4.73   Transmembrane  127 - 143 ( 124 - 144)
   INTEGRAL    Likelihood = -4.19   Transmembrane   50 -  66 (  47 -  67)
   PERIPHERAL  Likelihood =  3.76     201
 modified ALOM score:   2.82

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13178 GB:Z99110 ykoC [Bacillus subtilis]
 Identities = 61/226 (26%), Positives = 108/226 (46%), Gaps = 12/226 (5%)

Query: 49  FLIVVSLGSLVLFRLAKIKWQQVSFVMTLVVVFAVLNIIMVYLFAPHYGDKIYGSSSLLL 108
            F I++  G L+   +   KW      +   + F +L    V+  A     K+  +    L
Sbjct: 36  FYIIIVAGVLLAAGIPLKKW------LLFTIPFLILAFGCVWTAAVF--GKVPTTPDNFL 87

Query: 109 KGIGPYDVTSQELFYLFNLILKYFCTVPLALLFLMTTNPSQFASSL-NQLGLSYKIAYAV 167
               GP  + S  +    +L  +  C   L+++F+ TT+P  F  SL  Q  LS K+AY V
Sbjct: 88  FQAGPISINSDNVSVGISLGFRILCFSALSMMFVFTTDPILFMLSLVQQCRLSPKLAYGV 147

Query: 168 SLTLRYIPDVQEEFYTIRRAQEARGIELSKKSNLVARIKGNLQIVTPLIFSSLERIDTVA 227
                R++P +++E   I++A + RG   + +S ++ +I    +   PL+ S++ + + A
Sbjct: 148 IAGFRFLPLLKDEVQLIQQAHKIRGG--AAESGIINKISALKRYTIPLLASAIRKAERTA 205

Query: 228 TAMELRRFGKNKRRTWYSKQSLEKSDIVLIILALASLFVSLYLIHL 273
             AME + F  ++ RT+Y   S+ + D V   L L  LF   +L+ L
Sbjct: 206 LAMESKGFTGSRNRTYYRTLSVNRRDWVFFCLVLL-LFAGSFLVSL 250
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1577

A DNA sequence (GBSx1671) was identified in *S.agalactiae* <SEQ ID 4873> which encodes the amino acid sequence <SEQ ID 4874>. This protein is predicted to be cobalt ABC transporter, ATP-binding protein (cbiO). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.91   Transmembrane  436 - 452 ( 435 - 452)

----- Final Results -----
              bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13179 GB:Z99110 similar to cation ABC transporter
           (ATP-binding protein) [Bacillus subtilis]
 Identities = 151/483 (31%), Positives = 248/483 (51%), Gaps = 19/483 (3%)

Query: 8   KDFTFQYDVQSEPTLKGINLSIPKGEKVLILGPSGSGKSTLGHCLNGIIPNTHKGQYSGI 67
           +  +F Y+   +P  + I+  + KGE VL+LGPSG GKS+L  CLNG+ P    G  SG
Sbjct: 11  EQLSFSYEEDEKPVFQDISFELQKGECVLLLGPSGCGKSSLALCLNGLYPEACDGIQSGH 70

Query: 68  FTINHKNAFDLSIYDK-SHLVSTVLQDPDGQFIGLTVAEDIAFALENDVVAQEEMASIVE 126
             +  K   D    +  +     V QDPD QF  LTV ++IAF LEN  + +EEM   +
Sbjct: 71  VFLFQKPVTDAETSETITQHAGVVFQDPDQQFCMLTVEDEIAFGLENLQIPKEEMTEKIN 130

Query: 127 MWAKRLEIAPLLSKRPQDLSGGQKQRVSLAGVLVDDSPILLFDEPLANLDPQSGQDIMAL 186
               +L I  L  K    LSGGQKQ+V+LA +L  +  +++ DEP + LDP S ++ + L
Sbjct: 131 AVLGKLRITHLKEKMISTLSGGQKQKVALACILAMEPELIILDEPTSLLDPFSAREFVHL 190

Query: 187 VDRIHQEQDATTIIIEHRLED--VFYERVDRVVLFSDGQIIYNGEPDQLL--KTNFLSEY 242
           +  + +E+  + ++IEH+L++   + ER   +VL   G+   +G    L   +   L +
Sbjct: 191 MKDLQREKGFSLLVIEHQLDEWAPWIERT--IVLDKSGKKALDGLTKNLFQHEAETLKKL 248

Query: 243 GIREPLYISALKNLGYDFEKQNTMTSIDDFDFSELLIPKMRALDLDKHTDKLLSVQHLSV 302
           GI  P      + L   F     M   +        + K +A       + +L V  LS
Sbjct: 249 GIAIPKVCHLQEKLSMPFTLSKEMLFKEPIPAGH--VKKKKA----PSGESVLEVSSLSF 302

Query: 303 SYDLENNTLDDVSFDLYKGQRLAIVGKNGAGKSTLAKALCQFI-PNNATLIYNNEDVSQD 361
           +   +     D+SF L +G   A+VG NG GKSTL   L   + P +  ++  ++ + +
Sbjct: 303 ARG-QQAIFKDISFSLREGSLTALVGPNGTGKSTLLSVLASLMKPQSGKILLYDQPLQKY 361

Query: 362 SIKERAERIGYVLQNPNQMISQAMVFDEVALGLRLRGFSDNDIESRVYDILKVCGLYQFR 421
             KE  +R+G+V QNP        V+DE+  G +    ++ + E +   +L+  GL
Sbjct: 362 KEKELRKRMGFVFQNPEHQFVTDTVYDELLFGQK----ANAETEKKAQHLLQRFGLAHLA 417

Query: 422 NWPISALSFGQKKRVTIASILILNPEVIILDEPTAGQDMKHYTEMMSFLDKLSCDGHTIV 481
           +    A+S GQK+R+++A++L+ + +V++LDEPT GQD +   E M  + ++  +G  ++
Sbjct: 418 DHHPFAISQGQKRRLSVATMLMHDVKVLLLDEPTFGQDARTAAECMEMIQRIKAEGTAVL 477

Query: 482 MIT 484
           MIT
Sbjct: 478 MIT 480
```

There is also homology to SEQ ID 4416.

SEQ ID 4874 (GBS424d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 2 & 4; MW 77kDa) and in Figure 239 (lane 10; MW 77kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 5 & 7; MW 52kDa) and in Figure 182 (lane 4; MW 52kDa). Purified GBS424d-His is shown in Figure 241, lanes 6 & 7. Purified GBS424d-GST is shown in Figure 246, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1578

A DNA sequence (GBSx1672) was identified in *S.agalactiae* <SEQ ID 4875> which encodes the amino acid sequence <SEQ ID 4876>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12    Transmembrane   39 -  55 (  35 -  63)
   INTEGRAL    Likelihood = -3.98    Transmembrane   72 -  88 (  71 -  90)
   INTEGRAL    Likelihood = -3.66    Transmembrane  108 - 124 ( 106 - 127)
   INTEGRAL    Likelihood = -2.34    Transmembrane  182 - 198 ( 181 - 198)
   INTEGRAL    Likelihood = -1.44    Transmembrane  141 - 157 ( 139 - 158)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59830 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 109/182 (59%), Positives = 141/182 (76%)

Query: 31  MNTNTIKKVVATGIGAALFIIIGMLVNIPTPIPNTNIQLQYAVLALFAVIYGPGVGFFTG 90
           M  N++K VVATGIGAALF+IIG L+NIPTPIPNT+IQLQYAVLALF+ ++GP  GF  G
Sbjct: 1   MKNNSVKIVVATGIGAALFVIIGWLINIPTPIPNTSIQLQYAVLALFSALFGPLAGFLIG 60

Query: 91  FIGHALKDSIQYGSPWWTWVLVSGLLGLMIGFFAKKLAIQLSGMTKKDLLLFNVVQVIAN 150
           FIGHALKDS  YG+PWWTWVL SGL+GL +GF  K+ ++       K+++ FN+VQ +AN
Sbjct: 61  FIGHALKDSFLYGAPWWTWVLGSGLMGLFLGFGVKRESLTQGIFGNKEIIRFNIVQFLAN 120

Query: 151 LIGWSVVAPYGDIFFYSEPASKVFAQGFLSSLVNSITIGVGGTLLLLAYAKSRPQKGSLS 210
           ++ W ++AP GDI  YSEPA+KVF QG ++ LVN++TI V GTLLL  YA +R + G+L
Sbjct: 121 VVVWGLIAPIGDILVYSEPANKVFTQGVVAGLVNALTIAVAGTLLLKLYAATRTKSGTLD 180

Query: 211 KD 212
           K+
Sbjct: 181 KE 182
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8857> and protein <SEQ ID 8858> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     -5.01
GvH: Signal Score (-7.5): -5.9
     Possible site: 50
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 5 value:  -8.12 threshold:  0.0
   INTEGRAL    Likelihood = -8.12    Transmembrane   31 -  47 (  27 -  55)
   INTEGRAL    Likelihood = -3.98    Transmembrane   64 -  80 (  63 -  82)
   INTEGRAL    Likelihood = -3.66    Transmembrane  100 - 116 (  98 - 119)
   INTEGRAL    Likelihood = -2.34    Transmembrane  174 - 190 ( 173 - 190)
   INTEGRAL    Likelihood = -1.44    Transmembrane  133 - 149 ( 131 - 150)
   PERIPHERAL  Likelihood =  5.78       9
 modified ALOM score:   2.12
```

```
*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02330(367 - 912 of 1212)
GP|6165407|emb|CAB59830.1||AJ012388(1 - 182 of 182) hypothetical protein {Lactococcus
lactis}
%Match = 28.1
%Identity = 59.9  %Similarity = 78.6
Matches = 109  Mismatches = 39  Conservative Sub.s = 34

102       132       162       192       222       252       282       312
MQVVGVGFIVGVIQDSCETALNSSTDVLFTAVAEKSVFGKK*TNEGLRYSI*DLFWYLILFSIVFQFFLSIRFQISLKYD

342       372       402       432       462       492       522       552
KIEQIVSDCLSLFFREVFMNTNTIKKVVATGIGAALFIIIGMLVNIPTPIPNTNIQLQYAVLALFAVIYGPGVGFFTGFI
                  |  |::| |||||||||||:||| |:|||||||||:|||||||||||: ::||  ||: |||
                  MKNNSVKIVVATGIGAALFVIIGWLINIPTPIPNTSIQLQYAVLALFSALFGPLAGFLIGFI
                          10        20        30        40        50        60

582       612       642       672       702       732       762       792
GHALKDSIQYGSPWWTWVLVSGLLGLMIGFFAKKLAIQLSGMTKKDLLLFNVVQVIANLIGWSVVAPYGDIFFYSEPASK
|||||||  ||:||||||| |||:|| :||  |: ::       |::: ||:|| :||:: | ::|| |||: |||||:|
GHALKDSFLYGAPWWTWVLGSGLMGLFLGFGVKRESLTQGIFGNKEIIRFNIVQFLANVVVWGLIAPIGDILVYSEPANK
                80        90       100       110       120       130       140

822       852       882       912       942       972       1002      1032
VFAQGFLSSLVNSITIGVGGTLLLLAYAKSRPQKGSLSKD*DKRVIYERFY*MEGFYLSI*RSI*TNFKRD*LKHS*R*K
|| || :: |||::|| | |||||  || :| : |:| |:
VFTQGVVAGLVNALTIAVAGTLLLKLYAATRTKSGTLDKE
               160       170       180
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1579

A DNA sequence (GBSx1673) was identified in *S.agalactiae* <SEQ ID 4877> which encodes the amino acid sequence <SEQ ID 4878>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.85   Transmembrane   86 - 102 (  80 - 106)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1580

A DNA sequence (GBSx1674) was identified in *S.agalactiae* <SEQ ID 4879> which encodes the amino acid sequence <SEQ ID 4880>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.61   Transmembrane  107 - 123 (  96 - 124)
   INTEGRAL    Likelihood = -1.86   Transmembrane  124 - 140 ( 124 - 142)
   INTEGRAL    Likelihood = -1.38   Transmembrane   83 -  99 (  83 - 100)
   INTEGRAL    Likelihood = -1.12   Transmembrane  142 - 158 ( 142 - 160)

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9415> which encodes amino acid sequence <SEQ ID 9416> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76124 GB:AE000391 putative transport protein [Escherichia
           coli K12]
 Identities = 139/178 (78%), Positives = 159/178 (89%)

Query: 1   MVGTMLFVALVVNPIIAFVMMRKNPYPLVLRCLKDSGITAFFTRSSAANIPVNMRLCEDL 60
           +VG ML VALVVNP++ +  +R+NP+PLVL CL++SG+ AFFTRSSAANIPVNM LCE L
Sbjct: 222 LVGCMLLVALVVNPLLVWWKIRRNPFPLVLLCLRESGVYAFFTRSSAANIPVNMALCEKL 281

Query: 61  GLDKDTYSVSIPLGAAINMAGAAITINILTLAAVNTLGITVDFPTAFLLSVVAAVSACGA 120
            LD+DTYSVSIPLGA INMAGAAITI +LTLAAVNTLGI VD PTA LLSVVA++ ACGA
Sbjct: 282 NLDRDTYSVSIPLGATINMAGAAITITVLTLAAVNTLGIPVDLPTALLLSVVASLCACGA 341

Query: 121 SGVTGGSLLLIPVACSLFGISNDVAMQVVGVGFIVGVIQDSCETALNSSTDVLFTAVA 178
           SGV GGSLLLIP+AC++FGISND+AMQVV VGFI+GV+QDSCETALNSSTDVLFTA A
Sbjct: 342 SGVAGGSLLLIPLACNMFGISNDIAMQVVAVGFIIGVLQDSCETALNSSTDVLFTAAA 399
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4881> which encodes the amino acid sequence <SEQ ID 4882>. Analysis of this protein sequence reveals the following:

```
     Possible site: 58
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.69   Transmembrane  212 - 228 ( 202 - 239)
   INTEGRAL    Likelihood = -7.38   Transmembrane   78 -  94 (  74 - 108)
   INTEGRAL    Likelihood = -6.53   Transmembrane  179 - 195 ( 175 - 200)
   INTEGRAL    Likelihood = -6.10   Transmembrane  315 - 331 ( 312 - 341)
   INTEGRAL    Likelihood = -5.36   Transmembrane   44 -  60 (  42 -  61)
   INTEGRAL    Likelihood = -4.41   Transmembrane   13 -  29 (  11 -  41)
   INTEGRAL    Likelihood = -3.19   Transmembrane  340 - 356 ( 333 - 358)
   INTEGRAL    Likelihood = -3.08   Transmembrane  145 - 161 ( 144 - 162)
   INTEGRAL    Likelihood = -0.90   Transmembrane  358 - 374 ( 358 - 376)

----- Final Results -----
               bacterial membrane --- Certainty=0.6477(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF95950 GB:AE004347 sodium/dicarboxylate symporter [Vibrio cholerae]
 Identities = 243/385 (63%), Positives = 299/385 (77%), Gaps = 2/385 (0%)

Query: 9   VRVSLIKKIGIGVVIGVMLGILAPDLTG-FSILGKLFVGGLKAIAPLLVFALVSQAISHQ 67
           VR +L+ +I  G+++G  +   +P+      ++G LFVG LKA+AP+LVF LV+ +I++Q
```

```
Sbjct: 11  VRGNLVLQILAGILLGAAMATFSPEYAQKVGLIGNLFVGALKAVAPVLVFILVASSIANQ 70

Query: 68  KKGKQTNMTLIIVLYLFGTFASALVAVLTAYLFPLTLVLNTPVNTELSPPQGVAEVFQSL 127
            KK + T M  I+VLYLFGTF++AL AV+ ++LFP TLVL T      +PPQG+AEV  +L
Sbjct: 71  KKNQHTYMRPIVVLYLFGTFSAALTAVILSFLFPTTLVLATGAEGA-TPPQGIAEVLNTL 129

Query: 128 LLKLVDNPINALATANYIGVLSWAIIFGLALKAASKETKHLIKTAAEVTSQIVVWIINLA 187
            L KLVDNP++AL  ANYIG+L+W +  GLAL  +S  TK + +  +   SQIV +II LA
Sbjct: 130 LFKLVDNPVSALMNANYIGILAWGVGLGLALHHSSSTTKAVFEDLSHGISQIVRFIIRLA 189

Query: 188 PIGIMSLVFTTISENGVGILSDYAFLILVLVGTMLFVALVVNPLIAVLITRQNPYPLVLR 247
            P GI  LV +T +   G   L+ YA L+ VL+G M F+ALVVNP+I     R+NP+PLVL+
Sbjct: 190 PFGIFGLVASTFATTGFDALAGYAQLLAVLLGAMAFIALVVNPMIVYYKIRRNPFPLVLQ 249

Query: 248 CLRESGLTAFFTRSSAANIPVNMQLCQKIGLSKDTYSVSIPLGATINMGGAAITINVLTL 307
            CLRESG+TAFFTRSSAANIPVNM LC+K+ L +DTYSVSIPLGATINM GAAITI VLTL
Sbjct: 250 CLRESGVTAFFTRSSAANIPVNMALCEKLKLDEDTYSVSIPLGATINMAGAAITITVLTL 309

Query: 308 AAVHTFGIPIDFLTALLLSVVAAVSACGASGVAGGSLLLIPVACSLFGISNDLAMQVVGV 367
            AAVHT GI +D +TALLLSVVAAVSACGASGVAGGSLLLIP+AC LFGISND+AMQVV V
Sbjct: 310 AAVHTMGIEVDLMTALLLSVVAAVSACGASGVAGGSLLLIPLACGLFGISNDIAMQVVAV 369

Query: 368 GFIVGVIQDSCETALNSSTDVLFTA 392
            GFI+GVIQDS ETALNSSTDVLFTA
Sbjct: 370 GFIIGVIQDSAETALNSSTDVLFTA 394
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/186 (82%), Positives = 172/186 (92%)

Query: 1   MVGTMLFVALVVNPIIAFVMMRKNPYPLVLRCLKDSGITAFFTRSSAANIPVNMRLCEDL 60
            +VGTMLFVALVVNP+IA ++  R+NPYPLVLRCL++SG+TAFFTRSSAANIPVNM+LC+ +
Sbjct: 217 LVGTMLFVALVVNPLIAVLITRQNPYPLVLRCLRESGLTAFFTRSSAANIPVNMQLCQKI 276

Query: 61  GLDKDTYSVSIPLGAAINMAGAAITINILTLAAVNTLGITVDFPTAFLLSVVAAVSACGA 120
            GL KDTYSVSIPLGA INM GAAITIN+LTLAAV+T GI +DF TA LLSVVAAVSACGA
Sbjct: 277 GLSKDTYSVSIPLGATINMGGAAITINVLTLAAVHTFGIPIDFLTALLLSVVAAVSACGA 336

Query: 121 SGVTGGSLLLIPVACSLFGISNDVAMQVVGVGFIVGVIQDSCETALNSSTDVLFTAVAEK 180
            SGV GGSLLLIPVACSLFGISND+AMQVVGVGFIVGVIQDSCETALNSSTDVLFTA+AE
Sbjct: 337 SGVAGGSLLLIPVACSLFGISNDLAMQVVGVGFIVGVIQDSCETALNSSTDVLFTAIAEN 396

Query: 181 SVFGKK 186
            + + +K
Sbjct: 397 AFWKRK 402
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1581

A DNA sequence (GBSx1675) was identified in *S.agalactiae* <SEQ ID 4883> which encodes the amino acid sequence <SEQ ID 4884>. This protein is predicted to be acid phosphatase. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2436(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9427> which encodes amino acid sequence <SEQ ID 9428> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
 Identities = 167/251 (66%), Positives = 209/251 (82%)

Query: 7   EQKTKFKNISLSSNKLLAKENTMSVLWYQNSAEAKALYLQGYNVAKMKLDDWLQKPSEKP 66
            ++  K  ++  S  +L + ENTMSVLWYQ +AEAKALYLQGY +A  +L + L + ++KP
Sbjct: 34  KETVKQTKVTYSDEQLRSNENTMSVLWYQRAAEAKALYLQGYQLATDRLKNQLGQATDKP 93

Query: 67  YSIILDLDETVLDNSPYQAKNIKDGSSFTPESWDKWVQKKSAKAVAGAKEFLKYANEKGI 126
            YSI+LD+DETVLDNSPYQAKNI +G+SFTPESWD WVQKK AK VAGAKEFL++A++ G+
Sbjct: 94  YSIVLDIDETVLDNSPYQAKNILEGTSFTPESWDVWVQKKEAKPVAGAKEFLQFADQNGV 153

Query: 127 KIYYVSDRTDAQVDATKENLEKEGIPVQGKDHLLFLKKGMKSKESRRQAVQKDTNLIMLF 186
            +IYY+SDR  +QVDAT ENL+KEGIPVQG+DHLLFL++G+KSKE+RRQ V++ TNLIMLF
Sbjct: 154 QIYYISDRAVSQVDATMENLQKEGIPVQGRDHLLFLEEGVKSKEARRQKVKETTNLIMLF 213

Query: 187 GDNLVDFADFSKSSSTDREQLLTKLQSEFGSKFIVFPNPMYGSWESAIYQGKHLDVQKQL 246
            GDNLVDFADFSK S  DR  LL++LQ EFG +FI+FPNPMYGSWESA+Y+G  LD   QL
Sbjct: 214 GDNLVDFADFSKKSEEDRTALLSELQEEFGRQFIIFPNPMYGSWESAVYKGDKLDASHQL 273

Query: 247 KERQKMLHSYD 257
            KER+K L S++
Sbjct: 274 KERRKALESFE 284
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4885> which encodes the amino acid sequence <SEQ ID 4886>. Analysis of this protein sequence reveals the following:

```
      Possible site: 25

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
 Identities = 234/284 (82%), Positives = 261/284 (91%)

Query: 1   MKSKKVVSVISLTLSLFLVTGCAKVDNNKSVNLKPATKQTYNSYSDDQLRSRENTMSVLW 60
            MK+K+V SVISL LSLFLVTGCA++D+  +VN K   KQT  +YSD+QLRS ENTMSVLW
Sbjct: 1   MKTKQVASVISLALSLFLVTGCAQLDHKANVNSKETVKQTKVTYSDEQLRSNENTMSVLW 60

Query: 61  YQRAAETQALYLQGYQLATDRLKEQLNKPTDKPYSIVLDIDETVLDNSPYQAKNVLEGTG 120
            YQRAAE +ALYLQGYQLATDRLK QL + TDKPYSIVLDIDETVLDNSPYQAKN+LEGT
Sbjct: 61  YQRAAEAKALYLQGYQLATDRLKNQLGQATDKPYSIVLDIDETVLDNSPYQAKNILEGTS 120

Query: 121 FTPESWDYWVQKKEAKPVAGAKDFLQFADQNGVQIYYISDRSTTQVDATMENLQKEGIPV 180
            FTPESWD WVQKKEAKPVAGAK+FLQFADQNGVQIYYISDR+ +QVDATMENLQKEGIPV
Sbjct: 121 FTPESWDVWVQKKEAKPVAGAKEFLQFADQNGVQIYYISDRAVSQVDATMENLQKEGIPV 180

Query: 181 QGRDHLLFLEKGVKSKESRRQKVKETTNVTMLFGDNLLDFADFSKKSQEDRTALLSDLQE 240
            QGRDHLLFLE+GVKSKE+RRQKVKETTN+ MLFGDNL+DFADFSKKS+EDRTALLS+LQE
Sbjct: 181 QGRDHLLFLEEGVKSKEARRQKVKETTNLIMLFGDNLVDFADFSKKSEEDRTALLSELQE 240

Query: 241 EFGRRFIIFPNPMYGSWEGAIYKGEKLDVLKQLEERRKSLKSFK 284
            EFGR+FIIFPNPMYGSWE A+YKG+KLD   QL+ERRK+L+SF+
Sbjct: 241 EFGRQFIIFPNPMYGSWESAVYKGDKLDASHQLKERRKALESFE 284
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 166/247 (67%), Positives = 207/247 (83%)

Query: 10  TKFKNISLSSNKLLAKENTMSVLWYQNSAEAKALYLQGYNVAKMKLDDWLQKPSEKPYSI 69
            TK    S S ++L ++ENTMSVLWYQ +AE +ALYLQGY +A  +L + L KP++KPYSI
Sbjct: 37  TKQTYNSYSDDQLRSRENTMSVLWYQRAAETQALYLQGYQLATDRLKEQLNKPTDKPYSI 96

Query: 70  ILDLDETVLDNSPYQAKNIKDGSSFTPESWDKWVQKKSAKAVAGAKEFLKYANEKGIKIY 129
            +LD+DETVLDNSPYQAKN+ +G+ FTPESWD WVQKK AK VAGAK+FL++A++ G++IY
Sbjct: 97  VLDIDETVLDNSPYQAKNVLEGTGFTPESWDYWVQKKEAKPVAGAKDFLQFADQNGVQIY 156

Query: 130 YVSDRTDAQVDATKENLEKEGIPVQGKDHLLFLKKGMKSKESRRQAVQKDTNLIMLFGDN 189
            Y+SDR+  QVDAT ENL+KEGIPVQG+DHLLFL+KG+KSKESRRQ V++ TN+ MLFGDN
Sbjct: 157 YISDRSTTQVDATMENLQKEGIPVQGRDHLLFLEKGVKSKESRRQKVKETTNVTMLFGDN 216

Query: 190 LVDFADFSKSSSTDREQLLTKLQSEFGSKFIVFPNPMYGSWESAIYQGKHLDVQKQLKER 249
            L+DFADFSK S  DR  LL+ LQ EFG +FI+FPNPMYGSWE AIY+G+ LDV KQL+ER
Sbjct: 217 LLDFADFSKKSQEDRTALLSDLQEEFGRRFIIFPNPMYGSWEGAIYKGEKLDVLKQLEER 276

Query: 250 QKMLHSY 256
            +K L S+
Sbjct: 277 RKSLKSF 283
```

SEQ ID 9428 (GBS661) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 2 & 4; MW 61kDa + lane 3; MW 27kDa) and in Figure 186 (lane 11; MW 61kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 5-7; MW 25kDa).

GBS661-GST was purified as shown in Figure 237, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1582

A DNA sequence (GBSx1676) was identified in *S.agalactiae* <SEQ ID 4887> which encodes the amino acid sequence <SEQ ID 4888>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4889> which encodes the amino acid sequence <SEQ ID 4890>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 395/398 (99%), Positives = 398/398 (99%)

Query: 1   MAKLTVKDVDLKGKKVLVRVDFNVPLKDGVITNDNRITAALPTIKYIIEQGGRAILFSHL 60
```

```
           MAKLTVKDVDLKGKKVLVRVDFNVPLKDGVITNDNRITAALPTIKYIIEQGGRAILFSHL
Sbjct: 1   MAKLTVKDVDLKGKKVLVRVDFNVPLKDGVITNDNRITAALPTIKYIIEQGGRAILFSHL 60

Query: 61  GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPGVTRGAKLEEAINALEDGQVLLVENTRF 120
           GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPGVTRG+KLEEAINALEDGQVLLVENTRF
Sbjct: 61  GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPGVTRGSKLEEAINALEDGQVLLVENTRF 120

Query: 121 EDVDGKKESKNDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLLEN 180
           EDVDGKKESKNDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLLEN
Sbjct: 121 EDVDGKKESKNDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLLEN 180

Query: 181 EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENLLEKADKVLIGGGMTYTFYKAQGIEI 240
           EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENLLEKADKVLIGGGMTYTFYKAQGIEI
Sbjct: 181 EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENLLEKADKVLIGGGMTYTFYKAQGIEI 240

Query: 241 GNSLVEEDKLDVAKDLLEKSNGKLILPVDSKEANAFAGYTEVRDTEGEAVSEGFLGLDIG 300
           GNSLVEEDKLDVAKDLLEKSNGKLILPVDSKEANAFAGYTEVRDTEGEAVSEGFLGLDIG
Sbjct: 241 GNSLVEEDKLDVAKDLLEKSNGKLILPVDSKEANAFAGYTEVRDTEGEAVSEGFLGLDIG 300

Query: 301 PKSIAKFDEALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGGDSAAA 360
           PKSIA+FD+ALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGGDSAAA
Sbjct: 301 PKSIAEFDQALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGGDSAAA 360

Query: 361 AINLGRADKFSWISTGGGASMELLEGKVLPGLAALTEK 398
           AINLGRADKFSWISTGGGASMELLEGKVLPGLAALTEK
Sbjct: 361 AINLGRADKFSWISTGGGASMELLEGKVLPGLAALTEK 398
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1583

A DNA sequence (GBSx1677) was identified in *S.agalactiae* <SEQ ID 4891> which encodes the amino acid sequence <SEQ ID 4892>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.39   Transmembrane    97 - 113 (  93 - 118)
   INTEGRAL    Likelihood = -3.66   Transmembrane    25 -  41 (  24 -  48)
   INTEGRAL    Likelihood = -3.40   Transmembrane   121 - 137 ( 121 - 140)
   INTEGRAL    Likelihood = -3.24   Transmembrane    72 -  88 (  72 -  88)
   INTEGRAL    Likelihood = -2.07   Transmembrane   143 - 159 ( 143 - 160)

----- Final Results -----
               bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4893> which encodes the amino acid sequence <SEQ ID 4894>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.23   Transmembrane    97 - 113 (  93 - 118)
   INTEGRAL    Likelihood = -7.17   Transmembrane   121 - 137 ( 119 - 140)
   INTEGRAL    Likelihood = -4.19   Transmembrane    25 -  41 (  24 -  48)
   INTEGRAL    Likelihood = -3.24   Transmembrane    72 -  88 (  72 -  88)
   INTEGRAL    Likelihood = -2.55   Transmembrane   154 - 170 ( 154 - 170)

----- Final Results -----
               bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 155/178 (87%), Positives = 169/178 (94%)

Query: 1   MKTLKKLLSNYKFDIKKFKLGMRTFKTGLSVFLVLLVFHLFGWKGLQIGALTAVFSLRED 60
           MKTL+KLLSNYKFDIKKFKLGMRT KTGLSVFLVLLVFHLFGWKGLQIGALTAVFSLRED
Sbjct: 1   MKTLRKLLSNYKFDIKKFKLGMRTLKTGLSVFLVLLVFHLFGWKGLQIGALTAVFSLRED 60

Query: 61  FDKSVHFGFSRIIGNSIGGLLSLVFFAFNEIFHQAFWVTLLIVPICTMLCIMINVACNNK 120
           FDKSVHFGFSRIIGNSIGGLLSLVFFAFNEIFHQAFWVTLLIVPICTMLCIM+NVACNNK
Sbjct: 61  FDKSVHFGFSRIIGNSIGGLLSLVFFAFNEIFHQAFWVTLLIVPICTMLCIMVNVACNNK 120

Query: 121 SGIIGGTAALLIITLSIPSGETILYVFARIFETFCGVFIAMMVNTDIEILRKKLKNNK 178
           SGIIG  AALLIITLSIP+G+T +YV +R+FETFCGVF+A++VNTD+E+++ K  N K
Sbjct: 121 SGIIGAVAALLIITLSIPTGQTFIYVTSRVFETFCGVFVAILVNTDVELIKNKWFNKK 178
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1584

A DNA sequence (GBSx1678) was identified in *S.agalactiae* <SEQ ID 4895> which encodes the amino acid sequence <SEQ ID 4896>. This protein is predicted to be regulatory protein glnr (glnR). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
 Identities = 59/123 (47%), Positives = 89/123 (71%), Gaps = 5/123 (4%)

Query: 4   RELRRTMAVFPIGAVMKLTDLTARQIRYYEDQGLITPERTEGNRRMFSLNDMDRLLEIKD 63
           +E RR+  +FPIG VM LT L+ARQIRYYE+  L++P RT+GNRR+FS ND+D+LLEIKD
Sbjct: 2   KEDRRSAPLFPIGIVMDLTQLSARQIRYYEEHNLVSPTRTKGNRRLFSFNDVDKLLEIKD 61

Query: 64  FISDGLHISDIKNEYMQRQH-----KSKEKQKSLSDAEVRRLLQDELRNQGRFSSPSQHI 118
            +  GL+++ IK   + +++     K KE+ K +S  E+R++L+DEL++ GRF+  S
Sbjct: 62  LLDQGLNMAGIKQVLLMKENQTEAVKVKEETKEISKTELRKILRDELQHTGRFNRTSLRQ 121

Query: 119 GNM 121
           G++
Sbjct: 122 GDI 124
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4897> which encodes the amino acid sequence <SEQ ID 4898>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
 Identities = 59/122 (48%), Positives = 83/122 (67%), Gaps = 5/122 (4%)

Query: 4   KELRRSMAVFPIGTVMTLTDLSARQIRYYEDQGLIKPERTQGNRRMFSLNDMDRLLEIKD 63
           KE RRS  +FPIG VM LT LSARQIRYYE+  L+ P RT+GNRR+FS ND+D+LLEIKD
Sbjct: 2   KEDRRSAPLFPIGIVMDLTQLSARQIRYYEEHNLVSPTRTKGNRRLFSFNDVDKLLEIKD 61

Query: 64  FLSEGLNIAAIKREYVERQG-----KLMQKQKALTDADVRRILHDEMLTQSGFSTPSQHI 118
            L +GLN+A IK+  + ++      K+ ++ K ++  ++R+IL DE+     F+  S
Sbjct: 62  LLDQGLNMAGIKQVLLMKENQTEAVKVKEETKEISKTELRKILRDELQHTGRFNRTSLRQ 121

Query: 119 GN 120
           G+
Sbjct: 122 GD 123
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 90/123 (73%), Positives = 108/123 (87%)

Query: 1   MKERELRRTMAVFPIGAVMKLTDLTARQIRYYEDQGLITPERTEGNRRMFSLNDMDRLLE 60
           MKE+ELRR+MAVFPIG VM LTDL+ARQIRYYEDQGLI PERT+GNRRMFSLNDMDRLLE
Sbjct: 1   MKEKELRRSMAVFPIGTVMTLTDLSARQIRYYEDQGLIKPERTQGNRRMFSLNDMDRLLE 60

Query: 61  IKDFISDGLHISDIKNEYMQRQHKSKEKQKSLSDAEVRRLLQDELRNQGRFSSPSQHIGN 120
           IKDF+S+GL+I+ IK EY++RQ K  +KQK+L+DA+VRR+L DE+  Q  FS+PSQHIGN
Sbjct: 61  IKDFLSEGLNIAAIKREYVERQGKLMQKQKALTDADVRRILHDEMLTQSGFSTPSQHIGN 120

Query: 121 MHL 123
            +
Sbjct: 121 FRI 123
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1585

A DNA sequence (GBSx1679) was identified in *S.agalactiae* <SEQ ID 4899> which encodes the amino acid sequence <SEQ ID 4900>. This protein is predicted to be glutamine synthetase (glnA). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2157(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4901> which encodes the amino acid sequence <SEQ ID 4902>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00    Transmembrane  347 - 363 ( 347 - 363)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 392/448 (87%), Positives = 421/448 (93%)

Query: 1   MTITAEDIRREVKEKNVTFLRLMFTDILGVMKNVEIPATDEQLDKVLSNKAMFDGSSIEG 60
           M IT   DIRREVKEKNVTFLRLMFTDI+GVMKNVEIPAT EQLDKVLSNK MFDGSSIEG
Sbjct: 1   MAITVADIRREVKEKNVTFLRLMFTDIMGVMKNVEIPATKEQLDKVLSNKVMFDGSSIEG 60

Query: 61  FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEGEPFAGDPRGNLKRNMKRM 120
           FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEG+PFAGDPRGNLKR +K M
Sbjct: 61  FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEGKPFAGDPRGNLKRALKHM 120

Query: 121 QEMGYKSFNLGPEPEFFLFKMDENGNPTLDVNDKGGYFDLAPTDLADNTRREIVNVLTQM 180
            E+GYKSFNLGPEPEFFLFKMD+ GNPTL+VND GGYFDLAP DLADNTRREIVN+LT+M
Sbjct: 121 NEIGYKSFNLGPEPEFFLFKMDDKGNPTLEVNDNGGYFDLAPIDLADNTRREIVNILTKM 180

Query: 181 GFEVEASHHEVAVGQHEIDFKYDDVLKACDNIQLFKLVVKTIARKHGLYATFMAKPKFGI 240
           GFEVEASHHEVAVGQHEIDFKY DVLKACDNIQ+FKLVVKTIAR+HGLYATFMAKPKFGI
Sbjct: 181 GFEVEASHHEVAVGQHEIDFKYADVLKACDNIQIFKLVVKTIAREHGLYATFMAKPKFGI 240

Query: 241 NGSGMHCNMSLFDNEGNNAFFDPEDPRGMQLSEDAYYFLGGLMKHAYNYTAIINPTVNSY 300
            GSGMHCNMSLFDN+GNNAF+D  D RGMQLSEDAYYFLGGLMKHAYNYTAI NPTVNSY
Sbjct: 241 AGSGMHCNMSLFDNQGNNAFYDEADKRGMQLSEDAYYFLGGLMKHAYNYTAITNPTVNSY 300

Query: 301 KRLVPGYEAPVYVAWAGRNRSPLIRVPASRGMGTRLELRSVDPTANPYLALSVLLGSGLE 360
           KRLVPGYEAPVYVAWAG NRSPLIRVPASRGMGTRLELRSVDPTANPYLAL+VLL +GL+
Sbjct: 301 KRLVPGYEAPVYVAWAGSNRSPLIRVPASRGMGTRLELRSVDPTANPYLALAVLLEAGLD 360

Query: 361 GIENKIEAPEPIETNIYAMTVEERRQAGIVDLPSTLHNALEALEEDEVVKAALGTHIYTN 420
           GI NKIEAPEP+E NIY MT+EER +AGI+DLPSTLHNAL+AL++D+VV+ ALG HIYTN
Sbjct: 361 GIINKIEAPEPVEANIYTMTMEERNEAGIIDLPSTLHNALKALQKDDVVQKALGYHIYTN 420

Query: 421 FLDAKRIEWASYATYVSQWEIDNYLDLY 448
           FL+AKRIEW+SYAT+VSQWEID+Y+  Y
Sbjct: 421 FLEAKRIEWSSYATFVSQWEIDHYIHNY 448
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1586

A DNA sequence (GBSx1680) was identified in *S.agalactiae* <SEQ ID 4903> which encodes the amino acid sequence <SEQ ID 4904>. This protein is predicted to be SceB precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA66624 GB:X97985 ORF1 [Staphylococcus aureus]
 Identities = 44/119 (36%), Positives = 66/119 (54%), Gaps = 4/119 (3%)

Query: 26  SFASTNADANTYNYAVDVDYLASAEEIAQAHPA-SNTFPLGQCTWGVKE-MATWAGNWWG 83
           S AS +  +N  +        ++   I+  + + SN +  GQCT+ V + +    G+ WG
Sbjct: 117 SGASYSTTSNNVHVTTTAAPSSNGRSISNGYASGSNLYTSGQCTYYVFDRVGGKIGSTWG 176

Query: 84  NGGDWAASAASADYTVGTQPRVGSIVCWTDGSYGHVAYVTAVDPVTNKIQVLESNYAGH 142
           N  +WA +AAS+ YTV   P+VG+I+  T G YGHVAYV  V+     ++V E NY GH
Sbjct: 177 NASNWANAAASSGYTVNNTPKVGAIMQTTQGYYGHVAYVEGVNS-NGSVRVSEMNY-GH 233
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1013> which encodes the amino acid sequence <SEQ ID 1014>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 60/115 (52%), Positives = 81/115 (70%), Gaps = 7/115 (6%)

Query: 55  AHPASNTFPLGQCTWGVKEMATWAGNWWGNGGDWAASAASADYTVGTQPRVGSIVCWTDG 114
           ++ +SNT+P+GQCTWG K +A WAGN WGNGG WA SA +A Y  G+ P VG+I  W DG
Sbjct: 291 SYDSSNTYPVGQCTWGAKSLAPWAGNNWGNGGQWAYSAQAAGYRTGSTPMVGAIAVWNDG 350

Query: 115 SYGHVAYVTAVDPVTNKIQVLESNYAGHQWIDNYRGWFDPQNTVTPGVVSYIYPN 169
            YGHVA V  V   ++ I+V+ESNY+G Q+I ++RGWF+P        V++IYP+
Sbjct: 351 GYGHVAVVVEVQSASS-IRVMESNYSGRQYIADHRGWFNPTG------VTFIYPH 398
```

A related GBS gene <SEQ ID 8859> and protein <SEQ ID 8860> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      5.85
GvH: Signal Score (-7.5): 3.11
     Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   6.74 threshold:  0.0
   PERIPHERAL  Likelihood =  6.74     115
 modified ALOM score:  -1.85

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
37.5/56.7% over 200aa
                                                    Staphylococcus aureus
  GP|1340128| ORF1 Insert characterized

ORF00255(376 - 726 of 1107)
GP|1340128|emb|CAA66624.1||X97985(33 - 233 of 255) ORF1 {Staphylococcus aureus}
%Match = 9.0
%Identity = 37.5  %Similarity = 56.7
Matches = 45  Mismatches = 47  Conservative Sub.s = 23

294       324       354       384       414
SVIWI**TRSHQMEENMNIKQLKSKTMLGTVALVSAFSFASTNADANTYNYAVDVD--------------------~~~~
                  :  :         |  :   : | :| : :| | :|
     MKKIVTATIATAGLATIAFAGHDAQAAEQNNNGYNSNDAQSYSYTYTIDAQGNYHYTWTGNWNPSQLTQNN~~~~
             10        20        30        40        50        60        70

462       489       516       546       576       606
--------------------YLASAEEIAQAHPA-SNTFPLGQCTWGV-KEMATWAGNWWGNGGDWAASAASADYTVGTQ
                      ::   |:  : : || :  ||||: |   :    |: |||  :|| :|||: |||
GSGASYSTTSNNVHVTTTAAPSSNGRSISNGYASGSNLYTSGQCTYYVFDRVGGKIGSTWGNASNWANAAASSGYTVNNT
             130       140       150       160       170       180       190
```

```
636       666       696       726       756       786       816       846
PRVGSIVCWTDGSYGHVAYVTAVDPVTNKIQVLESNYAGHQWIDNYRGWFDPQNTVTPGVVSYIYPN*SIKNSSHRRYKS
|:||:|:  | | |||||||  |:     ::| | || ||                :   :
PKVGAIMQTTQGYYGHVAYVEGVNS-NGSVRVSEMNY-GHGAGVVTSRTISANQAGSYNFIH
            210       220        230        240       250
```

SEQ ID 8860 (GBS30) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 2; MW 19.2kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 2; MW 44.2kDa).

GBS30-GST was purified as shown in Figure 193, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1587

A DNA sequence (GBSx1681) was identified in *S.agalactiae* <SEQ ID 4905> which encodes the amino acid sequence <SEQ ID 4906>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.93    Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1588

A DNA sequence (GBSx1682) was identified in *S.agalactiae* <SEQ ID 4907> which encodes the amino acid sequence <SEQ ID 4908>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 353/550 (64%), Positives = 443/550 (80%)

Query: 6    LKPEEVGVYAIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYSYIVEN 65
            LK  +  VYA+GGLGEIGKNTY +++QDEII++DAGIKFPED+LLGIDYVIPDYSY+V+N
Sbjct: 4    LKNNQTAVYALGGLGEIGKNTYAVQFQDEIILIDAGIKFPEDELLGIDYVIPDYSYLVKN 63

Query: 66   IDRIKALVITHGHEDHIGGIPFLLKQANLPIYAGPLALALIKGKLEEHGLLRDATLYEIH 125
             ++IK L ITHGHEDHIGGIP+LL++ N+PIY G LAL L++GKLEEHGLLR A L++I
Sbjct: 64   ENKIKGLFITHGHEDHIGGIPYLLREVNIPIYGGKLALGLLRGKLEEHGLLRKAKLHDIQ 123
```

```
Query: 126 ANTELTFKNLSVTFFRTTHSIPEPLGIVIHTPQGKVICTGDFKFDFTPVGEPADLHRMAA 185
            +  + F   SV+FFRTTHSIP+  GIV+ TP G ++ TGDFKFDFTPVGEPA+L +MA
Sbjct: 124 EDDIIKFAKTSVSFFRTTHSIPDSYGIVVKTPPGNIVHTGDFKFDFTPVGEPANLTKMAK 183

Query: 186 LGEDGVLCLLSDSTNAEVPTFTNSEKIVGQSIMKIIEGIEGRIIFASFASNIFRLQQAAE 245
           +GE+GVLCLLSDSTN+E+P FT SE+ VG+SI  I   +EGRIIFA+FASNI RLQQA E
Sbjct: 184 IGEEGVLCLLSDSTNSEIPEFTMSERKVGESIDHIFRRVEGRIIFATFASNIHRLQQAVE 243

Query: 246 AAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFIEPSELKNLHASEVLIMCTGSQGE 305
           +AV+ GRK+AVFGRSME AI  G ELGYIK PK TFIEP++L  L  +EV+I+CTGSQGE
Sbjct: 244 SAVRYGRKVAVFGRSMESAINIGQELGYIKAPKNTFIEPNQLNKLPDNEVMILCTGSQGE 303

Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHGKINNI 365
            MAAL+R+A GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I
Sbjct: 304 PMAALSRVAFGTHRQIQIIPGDTVIFSSSPIPGNTLSVSKTINQLYKAGANVIHGSLNDI 363

Query: 366 HTSGHGGQQEQKLMLRLIKPKYFMPVHGEYRMQKVHAGLAVDTGIPKENIFIMENGDVLA 425
           HTSGHGGQ+EQKLMLRLIKPKYFMP+HGEYRM K+H  LA D G+P EN FIM+NGDVLA
Sbjct: 364 HTSGHGGQEEQKLMLRLIKPKYFMPIHGEYRMLKMHTKLAEDCGVPAENCFIMDNGDVLA 423

Query: 426 LTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRHDLSEDGVVLAVATVDFDSKMILAG 485
           L  D A IAG   +  +YVDGNGIGDIG  VLRDR  LSE+G+V+ V +++     + AG
Sbjct: 424 LHPDEAGIAGKIPSGSVYVDGNGIGDIGNIVLRDRRILSEEGLVVVVVSLNMKEYKVTAG 483

Query: 486 PDILSRGFIYMRESGDLIRESQHILFNAIRIALKNKDASIQSVNGAIVNALRPFLYEKTE 545
           PD++SRGF+YMRESGDLI+E+Q +L N ++  ++ K      +   I + L PFLY++T+
Sbjct: 484 PDLISRGFVYMRESGDLIQEAQRLLANHLQEVMERKTNQWSEIKNEITDVLGPFLYDRTK 543

Query: 546 REPIIIPMVL 555
           R+P+I+P+++
Sbjct: 544 RKPMILPIIM 553
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4909> which encodes the amino acid sequence <SEQ ID 4910>. Analysis of this protein sequence reveals the following:

```
        Possible site: 28

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane  468 - 484 ( 468 - 484)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 353/550 (64%), Positives = 444/550 (80%)

Query: 6   LKPNEVGVFAIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYSYIVDN 65
           LK N+  V+A+GGLGEIGKNTY +++QDEII++DAGIKFPED+LLGIDYVIPDYSY+V N
Sbjct: 4   LKNNQTAVYALGGLGEIGKNTYAVQFQDEIILIDAGIKFPEDELLGIDYVIPDYSYLVKN 63

Query: 66  LDRVKALVITHGHEDHIGGIPFLLKQANIPIYAGPLALALIRGKLEEHGLWREATVYEIN 125
            +++K L ITHGHEDHIGGIP+LL++ NIPIY G LAL L+RGKLEEHGL R+A +++I
Sbjct: 64  ENKIKGLFITHGHEDHIGGIPYLLREVNIPIYGGKLALGLLRGKLEEHGLLRKAKLHDIQ 123

Query: 126 HNTELTFKNMSVTFFKTTHSIPEPVGIVIHTPQGKIICTGDFKFDFTPVGDPADLQRMAA 185
            +  + F   SV+FF+TTHSIP+  GIV+ TP G I+ TGDFKFDFTPVG+PA+L +MA
Sbjct: 124 EDDIIKFAKTSVSFFRTTHSIPDSYGIVVKTPPGNIVHTGDFKFDFTPVGEPANLTKMAK 183

Query: 186 LGEEGVLCLLSDSTNAEIPTFTNSEKVVGQSILKIIEGIHGRIIFASFASNIYRLQQAAE 245
           +GEEGVLCLLSDSTN+EIP FT SE+ VG+SI  I   + GRIIFA+FASNI+RLQQA E
Sbjct: 184 IGEEGVLCLLSDSTNSEIPEFTMSERKVGESIDHIFRRVEGRIIFATFASNIHRLQQAVE 243
```

```
Query: 246 AAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFIEPSELKNLHASEVLIMCTGSQGE 305
           +AV+ GRK+AVFGRSME AI  G ELGYIK PK TFIEP++L  L  +EV+I+CTGSQGE
Sbjct: 244 SAVRYGRKVAVFGRSMESAINIGQELGYIKAPKNTFIEPNQLNKLPDNEVMILCTGSQGE 303

Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHGKVNNI 365
            MAAL+R+A GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I
Sbjct: 304 PMAALSRVAFGTHRQIQIIPGDTVIFSSSPIPGNTLSVSKTINQLYKAGANVIHGSLNDI 363

Query: 366 HTSGHGGQQEQKLMLSLIKPKYFMPVHGEYRMQKVHAGLAMDIGIPKENIFIMENGDVLA 425
           HTSGHGGQ+EQKLML LIKPKYFMP+HGEYRM K+H  LA D G+P EN FIM+NGDVLA
Sbjct: 364 HTSGHGGQEEQKLMLRLIKPKYFMPIHGEYRMLKMHTKLAEDCGVPAENCFIMDNGDVLA 423

Query: 426 LTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRRDLSEDGVVLAVATVDFNTQMILAG 485
           L  D A IAG   +  +YVDGNGIGDIG  VLRDRR LSE+G+V+ V +++     + AG
Sbjct: 424 LHPDEAGIAGKIPSGSVYVDGNGIGDIGNIVLRDRRILSEEGLVVVVVSLNMKEYKVTAG 483

Query: 486 PDILSRGFIYMRESGDLIRESQRVLFNAIRIALKNKDASIQSVNGAIVNALRPFLYEKTE 545
           PD++SRGF+YMRESGDLI+E+QR+L N ++  ++ K      +   I + L PFLY++T+
Sbjct: 484 PDLISRGFVYMRESGDLIQEAQRLLANHLQEVMERKTNQWSEIKNEITDVLGPFLYDRTK 543

Query: 546 REPIIIPMVL 555
           R+P+I+P+++
Sbjct: 544 RKPMILPIIM 553
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 523/559 (93%), Positives = 550/559 (97%)

Query: 1   MSNINLKPEEVGVYAIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS 60
           M+NI+LKP EVGV+AIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS
Sbjct: 1   MTNISLKPNEVGVFAIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS 60

Query: 61  YIVENIDRIKALVITHGHEDHIGGIPFLLKQANLPIYAGPLALALIKGKLEEHGLLRDAT 120
           YIV+N+DR+KALVITHGHEDHIGGIPFLLKQAN+PIYAGPLALALI+GKLEEHGL R+AT
Sbjct: 61  YIVDNLDRVKALVITHGHEDHIGGIPFLLKQANIPIYAGPLALALIRGKLEEHGLWREAT 120

Query: 121 LYEIHANTELTFKNLSVTFFRTTHSIPEPLGIVIHTPQGKVICTGDFKFDFTPVGEPADL 180
           +YEI+ NTELTFKN+SVTFF+TTHSIPEP+GIVIHTPQGK+ICTGDFKFDFTPVG+PADL
Sbjct: 121 VYEINHNTELTFKNMSVTFFKTTHSIPEPVGIVIHTPQGKIICTGDFKFDFTPVGDPADL 180

Query: 181 HRMAALGEDGVLCLLSDSTNAEVPTFTNSEKIVGQSIMKIIEGIEGRIIFASFASNIFRL 240
            RMAALGE+GVLCLLSDSTNAE+PTFTNSEK+VGQSI+KIIEGI GRIIFASFASNI+RL
Sbjct: 181 QRMAALGEEGVLCLLSDSTNAEIPTFTNSEKVVGQSILKIIEGIHGRIIFASFASNIYRL 240

Query: 241 QQAAEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFIEPSELKNLHASEVLIMCT 300
           QQAAEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFIEPSELKNLHASEVLIMCT
Sbjct: 241 QQAAEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFIEPSELKNLHASEVLIMCT 300

Query: 301 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG 360
           GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG
Sbjct: 301 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG 360

Query: 361 KINNIHTSGHGGQQEQKLMLRLIKPKYFMPVHGEYRMQKVHAGLAVDTGIPKENIFIMEN 420
           K+NNIHTSGHGGQQEQKLML LIKPKYFMPVHGEYRMQKVHAGLA+D GIPKENIFIMEN
Sbjct: 361 KVNNIHTSGHGGQQEQKLMLSLIKPKYFMPVHGEYRMQKVHAGLAMDIGIPKENIFIMEN 420

Query: 421 GDVLALTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRHDLSEDGVVLAVATVDFDSK 480
           GDVLALTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDR DLSEDGVVLAVATVDF+++
Sbjct: 421 GDVLALTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRRDLSEDGVVLAVATVDFNTQ 480

Query: 481 MILAGPDILSRGFIYMRESGDLIRESQHILFNAIRIALKNKDASIQSVNGAIVNALRPFL 540
           MILAGPDILSRGFIYMRESGDLIRESQ +LFNAIRIALKNKDASIQSVNGAIVNALRPFL
Sbjct: 481 MILAGPDILSRGFIYMRESGDLIRESQRVLFNAIRIALKNKDASIQSVNGAIVNALRPFL 540

Query: 541 YEKTEREPIIIPMVLTPDK 559
           YEKTEREPIIIPMVLTPDK
Sbjct: 541 YEKTEREPIIIPMVLTPDK 559
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1589

A DNA sequence (GBSx1683) was identified in *S.agalactiae* <SEQ ID 4911> which encodes the amino acid sequence <SEQ ID 4912>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2932(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13327 GB:Z99111 ykzG [Bacillus subtilis]
 Identities = 27/75 (36%), Positives = 44/75 (58%), Gaps = 7/75 (9%)

Query: 1  MIYKVFYQETKERNPRREQTKTLYVTIDAANELEGRIAARKLVEENTAYNIEFIELLSDK 60
          MIYKVFYQE  +  P RE+T +LY+   +  ++  ++  +K       +NIEFI  +
Sbjct: 1  MIYKVFYQEKADEVPVREKTDSLYIEGVSERDVRTKLKEKK-------FNIEFITPVDGA 53

Query: 61 HLEYEKETGVFELTE 75
           LEYE+++  F++ E
Sbjct: 54 FLEYEQQSENFKVLE 68
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4913> which encodes the amino acid sequence <SEQ ID 4914>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3428(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 60/76 (78%), Positives = 70/76 (91%)

Query: 1  MIYKVFYQETKERNPRREQTKTLYVTIDAANELEGRIAARKLVEENTAYNIEFIELLSDK 60
          MIYKVFYQETK+++PRRE TK LY+ IDA +EL+GRI AR+LVE+NT YN+EFIELLSDK
Sbjct: 1  MIYKVFYQETKDQSPRRESTKALYLNIDATDELDGRIKARRLVEDNTYYNVEFIELLSDK 60

Query: 61 HLEYEKETGVFELTEF 76
          HL+YEKETGVFELTEF
Sbjct: 61 HLDYEKETGVFELTEF 76
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1590

A DNA sequence (GBSx1684) was identified in *S.agalactiae* <SEQ ID 4915> which encodes the amino acid sequence <SEQ ID 4916>. This protein is predicted to be glycoprotein endopeptidase. Analysis of this protein sequence reveals the following:

```
Possible site: 13
```

```
>>> Seems to have no N-terminal signal sequence (or aa 1-17)

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0430(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA76861 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 94/182 (51%), Positives = 127/182 (69%), Gaps = 6/182 (3%)

Query: 2   MKVLAFDTSSKALSVAVLNNMECLATVTINIKKNHSINLMPAIDFLMQSIDLEPQDLDRI 61
           +++LA DTS++ LS+AV  N + L + T  +K+NHS+ LMPAID+LM  ++L P  +DR
Sbjct: 13  VRILAIDTSNQTLSIAVCENQKILGSYTATVKRNHSLTLMPAIDYLMSQLNLAPTAIDRF 72

Query: 62  VVAEGPGSYTGLRVAVATAKMLAYTLKIDLVGVSSLYAL-TNGFSENDLLVPLIDARRNN 120
           VVAEGPGSYTGLR+ V TAK LAYTLK +LVG+SSL AL  N   +  L+VPL DARR N
Sbjct: 73  VVAEGPGSYTGLRLGVTTAKTLAYTLKKELVGISSLQALAANCVGQTGLIVPLFDARRKN 132

Query: 121 VYVGFYQNGDTV----KPDCHTSLEEVLQEVGNKANVHFVGE-VAAFFDQIKKALPHAKI 175
           VY G Y+  D V     PD H SL E+L+++ N+ N+ FVGE V  F ++I + +PH +I
Sbjct: 133 VYAGAYRFVDGVWQNELPDQHISLRELLEQLKNEPNLFFVGEDVEKFTEEIAQIIPHGEI 192

Query: 176 TE 177
            +
Sbjct: 193 CD 194
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4917> which encodes the amino acid sequence <SEQ ID 4918>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38    Transmembrane   99 - 115 (  99 - 115)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9159> which encodes the amino acid sequence <SEQ ID 9160>. Analysis of this protein sequence reveals the following:

```
     Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38    Transmembrane   88 - 104 (  88 - 104)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 134/232 (57%), Positives = 172/232 (73%), Gaps = 3/232 (1%)

Query: 2   MKVLAFDTSSKALSVAVLNNMECLATVTINIKKNHSINLMPAIDFLMQSIDLEPQDLDRI 61
           MK LAFDTS+K LS+A+L++   LA +T+NI+K HS++LMPAIDFLM   DL+PQDL+RI
Sbjct: 12  MKTLAFDTSNKTLSLAILDDETLLADMTLNIQKKHSVSLMPAIDFLMTCTDLKPQDLERI 71

Query: 62  VVAEGPGSYTGLRVAVATAKMLAYTLKIDLVGVSSLYALTNGFSE---NDLLVPLIDARR 118
           VVA+GPGSYTGLRVAVATAK LAY+L I LVG+SSLYAL     +   N L+VPLIDARR
Sbjct: 72  VVAKGPGSYTGLRVAVATAKTLAYSLNIALVGISSLYALAASTCKQYPNTLVVPLIDARR 131

Query: 119 NNVYVGFYQNGDTVKPDCHTSLEEVLQEVGNKANVHFVGEVAAFFDQIKKALPHAKITET 178
            N YVG+Y+ G +V P  H SLE +++++  +  + FVGE A F ++I+K LP A +  T
```

```
Sbjct: 132 QNAYVGYYRQGKSVMPQAHASLEVIIEQLVEEGQLIFVGETAPFAEKIQKKLPQAILLPT 191

Query: 179 LPCAVAIGRKGQKMKSVNVDAFVPRYLKRVEAEENWLKNHCETNTEEYIKRV 230
           LP A   G  GQ +   NVDAFVP+YLKRVEAEENWLK++  +    Y+KR+
Sbjct: 192 LPSAYECGLLGQSLAPENVDAFVPQYLKRVEAEENWLKDNEIKDDSHYVKRI 243
```

SEQ ID 4916 (GBS69) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 9; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 4; MW 53.9kDa).

The GBS69-GST fusion product was purified (Figure 197, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 285), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1591

A DNA sequence (GBSx1685) was identified in *S.agalactiae* <SEQ ID 4919> which encodes the amino acid sequence <SEQ ID 4920>. This protein is predicted to be ribosomal-protein-alanine acetyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10137> which encodes amino acid sequence <SEQ ID 10138> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC06803 GB:AE000696 ribosomal-protein-alanine acetyltransferase
          [Aquifex aeolicus]
 Identities = 44/141 (31%), Positives = 74/141 (52%), Gaps = 8/141 (5%)

Query: 9   LREFEMESSEQALAIWSVLSDVYDKSPWSLSQISEDLKKDSTDYFFVYNDGEVIGFLALQ 68
           +RE E E  E+   ++ +  + +    WS     +D +   +  F +   DG+V+G++
Sbjct: 4   VREMEREDVER---VYEINRESFTTDAWSRFSFEKDFENKFSRRFVLEEDGKVVGYVIFW 60

Query: 69  QLVGEVEITNIAVKKNYQGKGYAYQLM----SMIADIEVPVFLEVRYSNIVAQKLYERCG 124
            +  E  I   A+   Y+GKGY  +L+    S + D    V L+VR SN+ A  LY++ G
Sbjct: 61  VVKEEATIMTFAIAPGYRGKGYGEKLLREAISRLGDKVKRVVLDVRKSNLRAINLYKKLG 120

Query: 125 FVVLRKRKNYYHDPIEDAIVM 145
           F V+ +RK YY D  E+A++M
Sbjct: 121 FKVVTERKGYYSDG-ENALLM 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4921> which encodes the amino acid sequence <SEQ ID 4922>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3800(Affirmative) < succ>
```

```
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 65/140 (46%), Positives = 96/140 (68%), Gaps = 1/140 (0%)

Query: 9   LREFEMES-SEQALAIWSVLSDVYDKSPWSLSQISEDLKKDSTDYFFVYNDGEVIGFLAL 67
           L E  M++  EQA  I+ +L  VY  SPW+L Q+  D+++D TDYF +Y+  +++GFLA+
Sbjct: 6   LSESNMKTVEEQAKNIYQLLEMVYGTSPWTLEQVLIDIRRDQTDYFLLYDHDKLLGFLAI 65

Query: 68  QQLVGEVEITNIAVKKNYQGKGYAYQLMSMIADIEVPVFLEVRYSNIVAQKLYERCGFVV 127
           Q L GEVE+T IA+  ++Q  G A QLM+ +  IE  +FLEVR SN  AQ LY++ GF
Sbjct: 66  QDLAGEVEMTQIAILPSHQELGLASQLMTHLDSIESDIFLEVRESNHRAQGLYQKFGFKF 125

Query: 128 LRKRKNYYHDPIEDAIVMRK 147
           + KR +YY +PIE A++M++
Sbjct: 126 IGKRPDYYRNPIETALLMKR 145
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1592

A DNA sequence (GBSx1686) was identified in *S.agalactiae* <SEQ ID 4923> which encodes the amino acid sequence <SEQ ID 4924>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0334(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1593

A DNA sequence (GBSx1687) was identified in *S.agalactiae* <SEQ ID 4925> which encodes the amino acid sequence <SEQ ID 4926>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.75    Transmembrane   86 - 102 (  86 - 104)

----- Final Results -----
               bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]
 Identities = 194/331 (58%), Positives = 263/331 (78%), Gaps = 1/331 (0%)

Query: 6   ILAVESSCDETSVAILKNDKELLANIIASQVESHKRFGGVVPEVASRHHVEVVTTCFEDA 65
           ILA+E+SCDETS A+++N   +L+N+++SQ++SHKRFGGVVPE+ASRHHVE +T   E+A
```

```
Sbjct: 12  ILAIETSCDETSAAVIENGTTILSNVVSSQIDSHKRFGGVVPEIASRHHVEQITVIVEEA 71

Query: 66  LQEAGIVASDLDAVAVTYGPGLVGALLVGMAAAKAFAWANKLPLIPINHMAGHLMAARDV 125
           + EAG+  +DL AVAVT GPGLVGALL+G+ AAKA A+A++LPLI ++H+AGH+ A R +
Sbjct: 72  MHEAGVDFADLAAVAVTEGPGLVGALLIGVNAAKAIAFAHQLPLIGVHHIAGHIYANRLL 131

Query: 126 KELQYPLLALLVSGGHTELVYVSEPGDYKIVGETRDDAVGEAYDKVGRVMGLTYPAGREI 185
           KEL++PLLAL+VSGGHTEL+Y+   G+++++GETRDDAVGEAYDKV R +GL YP G  I
Sbjct: 132 KELEFPLLALVVSGGHTELIYMENHGEFEVIGETRDDAVGEAYDKVARTLGLPYPGGPHI 191

Query: 186 DQLAHKGQDTYHFPRAMIKEDHLEFSFSGLKSAFINLHHNAEQKGEALVLEDLCASFQAA 245
           D+LA  G+DT  FPRA ++ D  +FSFSGLKSA IN  HNA+Q+GE +  ED+ ASFQA+
Sbjct: 192 DRLAVNGEDTLQFPRAWLEPDSFDFSFSGLKSAVINTLHNAKQRGENVQAEDVAASFQAS 251

Query: 246 VLDILLAKTQKALLKYPVKTLVVAGGVAANQGLRERLATDISPD-IDVVIPPLRLCGDNA 304
           V+D+L+ KT+KA  +Y V+ +++AGGVAAN+GLR  L      + ID+VIPPL LC DNA
Sbjct: 252 VIDVLVTKTKKAAEEYKVRQVLLAGGVAANKGLRTALEEAFFKEPIDLVIPPLSLCTDNA 311

Query: 305 GMIALAAAIEFEKENFASLKLNAKPSLAFES 335
            MI  AA+I+F+++ FA + LN +PSL  E+
Sbjct: 312 AMIGAAASIKFKQQTFAGMDLNGQPSLELEN 342
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4927> which encodes the amino acid sequence <SEQ ID 4928>. Analysis of this protein sequence reveals the following:

```
        Possible site: 38
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.76   Transmembrane   86 - 102 (  85 - 104)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]
 Identities = 196/330 (59%), Positives = 255/330 (76%), Gaps = 2/330 (0%)

Query: 6    ILAVESSCDETSVAILKNESTLLSNVIASQVESHKRFGGVVPEVASRHHVEVITTCFEDA 65
            ILA+E+SCDETS A+++N +T+LSNV++SQ++SHKRFGGVVPE+ASRHHVE IT   E+A
Sbjct: 12   ILAIETSCDETSAAVIENGTTILSNVVSSQIDSHKRFGGVVPEIASRHHVEQITVIVEEA 71

Query: 66   LQEAGISASDLSAVAVTYGPGLVGALLVGLAAAKAFAWANHLPLIPVNHMAGHLMAAREQ 125
            + EAG+  +DL+AVAVT GPGLVGALL+G+ AAKA A+A+ LPLI V+H+AGH+ A R
Sbjct: 72   MHEAGVDFADLAAVAVTEGPGLVGALLIGVNAAKAIAFAHQLPLIGVHHIAGHIYANRLL 131

Query: 126  KPLVYPLIALLVSGGHTELVYVPEPGDYHIIGETRDDAVGEAYDKVGRVMGLTYPAGREI 185
            K L +PL+AL+VSGGHTEL+Y+   G++ +IGETRDDAVGEAYDKV R +GL YP G  I
Sbjct: 132  KELEFPLLALVVSGGHTELIYMENHGEFEVIGETRDDAVGEAYDKVARTLGLPYPGGPHI 191

Query: 186  DQLAHKGQDTYHFPRAMITEDHLEFSFSGLKSAFINLHHNAKQKGDELILEDLCASFQAA 245
            D+LA  G+DT  FPRA +  D  +FSFSGLKSA IN  HNAKQ+G+ +  ED+ ASFQA+
Sbjct: 192  DRLAVNGEDTLQFPRAWLEPDSFDFSFSGLKSAVINTLHNAKQRGENVQAEDVAASFQAS 251

Query: 246  VLDILLAKTKKALSRYPAKMLVVAGGVAANQGLRDRLAQEI--THIEVVIPKLRLCGDNA 303
            V+D+L+ KTKKA   Y  + +++AGGVAAN+GLR  L +      I++VIP L LC DNA
Sbjct: 252  VIDVLVTKTKKAAEEYKVRQVLLAGGVAANKGLRTALEEAFFKEPIDLVIPPLSLCTDNA 311

Query: 304  GMIALAAAIEYDKQHFANMSLNAKPSLAFD 333
             MI  AA+I++ +Q FA M LN +PSL  +
Sbjct: 312  AMIGAAASIKFKQQTFAGMDLNGQPSLELE 341
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 288/334 (86%), Positives = 313/334 (93%), Gaps = 1/334 (0%)
```

```
Query: 1   MKDRYILAVESSCDETSVAILKNDKELLANIIASQVESHKRFGGVVPEVASRHHVEVVTT 60
           M DRYILAVESSCDETSVAILKN+  LL+N+IASQVESHKRFGGVVPEVASRHHVEV+TT
Sbjct: 1   MTDRYILAVESSCDETSVAILKNESTLLSNVIASQVESHKRFGGVVPEVASRHHVEVITT 60

Query: 61  CFEDALQEAGIVASDLDAVAVTYGPGLVGALLVGMAAAKAFAWANKLPLIPINHMAGHLM 120
           CFEDALQEAGI ASDL AVAVTYGPGLVGALLVG+AAAKAFAWAN LPLIP+NHMAGHLM
Sbjct: 61  CFEDALQEAGISASDLSAVAVTYGPGLVGALLVGLAAAKAFAWANHLPLIPVNHMAGHLM 120

Query: 121 AARDVKELQYPLLALLVSGGHTELVYVSEPGDYKIVGETRDDAVGEAYDKVGRVMGLTYP 180
           AAR+ K L YPL+ALLVSGGHTELVYV EPGDY I+GETRDDAVGEAYDKVGRVMGLTYP
Sbjct: 121 AAREQKPLVYPLIALLVSGGHTELVYVPEPGDYHIIGETRDDAVGEAYDKVGRVMGLTYP 180

Query: 181 AGREIDQLAHKGQDTYHFPRAMIKEDHLEFSFSGLKSAFINLHHNAEQKGEALVLEDLCA 240
           AGREIDQLAHKGQDTYHFPRAMI EDHLEFSFSGLKSAFINLHHNA+QKG+ L+LEDLCA
Sbjct: 181 AGREIDQLAHKGQDTYHFPRAMITEDHLEFSFSGLKSAFINLHHNAKQKGDELILEDLCA 240

Query: 241 SFQAAVLDILLAKTQKALLKYPVKTLVVAGGVAANQGLRERLATDISPDIDVVIPPLRLC 300
           SFQAAVLDILLAKT+KAL +YP K LVVAGGVAANQGLR+RLA +I+  I+VVIP LRLC
Sbjct: 241 SFQAAVLDILLAKTKKALSRYPAKMLVVAGGVAANQGLRDRLAQEIT-HIEVVIPKLRLC 299

Query: 301 GDNAGMIALAAAIEFEKENFASLKLNAKPSLAFE 334
           GDNAGMIALAAAIE++K++FA++ LNAKPSLAF+
Sbjct: 300 GDNAGMIALAAAIEYDKQHFANMSLNAKPSLAFD 333
```

SEQ ID 4926 (GBS371) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 7; MW 41kDa), in Figure 170 (lane 4 & 5; MW 55kDa) and in Figure 239 (lane 6; MW 55kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 7; MW 65kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1594

A DNA sequence (GBSx1688) was identified in *S.agalactiae* <SEQ ID 4929> which encodes the amino acid sequence <SEQ ID 4930>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1027(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1595

A DNA sequence (GBSx1689) was identified in *S.agalactiae* <SEQ ID 4931> which encodes the amino acid sequence <SEQ ID 4932>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1307(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1596

A DNA sequence (GBSx1690) was identified in *S.agalactiae* <SEQ ID 4933> which encodes the amino acid sequence <SEQ ID 4934>. This protein is predicted to be L4171-60 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10135> which encodes amino acid sequence <SEQ ID 10136> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24656 GB:AE001274 L4171.5 [Leishmania major]
 Identities = 118/282 (41%), Positives = 167/282 (58%), Gaps = 4/282 (1%)

Query: 2   GGTQTNQVVISSMLASYEGVIAAETGHVSSHEAGAIEFSGHKVLTLPSHNGKLLASEVAT 61
            GGTQTN +  S  L  +E VIA + GH+S+HE GAIE +GHKV+T P  +GKL  ++
Sbjct: 74  GGTQTNLIACSLALRPWEAVIATQLGHISTHETGAIEATGHKVVTAPCPDGKLRVAD--- 130

Query: 62  YIETFYADGNYQHMVFPGMVYISHPTEYGTLYSKAELEELSKICKHYQIPLFIDGARLGY 121
             IE+   +   +HMV P +VYIS+ TE GT Y+K ELE++S  CK + + LF+DGARL
Sbjct: 131 -IESALHENRSEHMVIPKLVYISNTTEVGTQYTKQELEDISASCKEHGLYLFLDGARLAS 189

Query: 122 GLAAKDTDVDFPTIAALSDVFYIGGTKMGALAGEAVVFTKKNRPKQFTTIVKQHGALLAK 181
             L++   D+    IA L+D+FYIG TK G + GEA++            ++KQ GAL+AK
Sbjct: 190 ALSSPVNDLTLADIARLTDMFYIGATKAGGMFGEALIILNDALKPNARHLIKQRGALMAK 249

Query: 182 GRLLGLAFDRFFTDNLYLKIGKHAIDLAEELKIILEEKGYSFYLKSPTNQQFIIVENTKL 241
            G LLG+ F+    DNL+ ++G H+  +A  LK  LE  G      S +NQ F I+ENT +
Sbjct: 250 GWLLGIQFEVLMKDNLFFELGAHSNKMAAILKAGLEACGIRLAWPSASNQLFPILENTMI 309

Query: 242 ADLAKNVAYSFWEKYDDHHTVIRLATSWSTSREDVTALRNVL 283
            A+L  +      E   D   ++RL TSW+T  ++      VL
Sbjct: 310 AELNNDFDMYTVEPLKDGTCIMRLCTSWATEEKECHRFVEVL 351
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 4934 (GBS648) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 8-10; MW 60kDa) and in Figure 186 (lane 6; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 12; MW 35kDa), in Figure 140 (lane 10; MW 35kDa) and in Figure 178 (lane 7; MW 35kDa).

Purified GBS648-GST is shown in Figure 243, lane 6; purified GBS648-His is shown in Fig. 229, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1597

A DNA sequence (GBSx1691) was identified in *S.agalactiae* <SEQ ID 4935> which encodes the amino acid sequence <SEQ ID 4936>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1598

A DNA sequence (GBSx1692) was identified in *S.agalactiae* <SEQ ID 4937> which encodes the amino acid sequence <SEQ ID 4938>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3848(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12716 GB:Z99108 similar to ribosomal protein S14 [Bacillus subtilis]
 Identities = 67/89 (75%), Positives = 76/89 (85%)

Query: 1  MAKKSKIAKFQKQQKLVEQYAELRRELKEKGDYEALRKLPKDSNPNRLKNRDLIDGRPHA 60
          MAKKSK+AK  K+Q+LVEQYA +RRELKEKGDYEAL KLP+DS P RL NR ++ GRP A
Sbjct: 1  MAKKSKVAKELKRQQLVEQYAGIRRELKEKGDYEALSKLPRDSAPGRLHNRCMVTGRPRA 60

Query: 61 YMRKFGMSRINFRNLAYKGQIPGIKKASW 89
          YMRKF MSRI FR LA+KGQIPG+KKASW
Sbjct: 61 YMRKFKMSRIAFRELAHKGQIPGVKKASW 89
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4939> which encodes the amino acid sequence <SEQ ID 4940>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 73/89 (82%), Positives = 85/89 (95%)

Query: 1  MAKKSKIAKFQKQQKLVEQYAELRRELKEKGDYEALRKLPKDSNPNRLKNRDLIDGRPHA 60
           MAKKSKIAK+QKQ +L+EQYA+LRR+LK KGDYE+LRKLP+DSNPNRLKNRD IDGRPHA
Sbjct: 1  MAKKSKIAKYQKQLQLIEQYADLRRDLKAKGDYESLRKLPRDSNPNRLKNRDKIDGRPHA 60

Query: 61 YMRKFGMSRINFRNLAYKGQIPGIKKASW 89
           YMRKFG+SRINFR+LA+KGQ+PG+ KASW
Sbjct: 61 YMRKFGVSRINFRDLAHKGQLPGVTKASW 89
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1599

A DNA sequence (GBSx1693) was identified in *S.agalactiae* <SEQ ID 4941> which encodes the amino acid sequence <SEQ ID 4942>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5183(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB95931 GB:AL359989 galactose-1-phosphate uridylyltransferase
          [Streptomyces coelicolor A3(2)]
 Identities = 31/105 (29%), Positives = 51/105 (48%), Gaps = 4/105 (3%)

Query: 27  DKCPFC--DKSQLGKILDVKDDMIWVENKYPTL--EETYQTLVIESNDHNGDISVYSESK 82
            D+CP C  D  +L +I D   D++  EN++P+L  +     +V  ++DH+   +  SE +
Sbjct: 68  DQCPLCPSDGERLSEIPDSAYDVVVFENRFPSLAGDSGRCEVVCFTSDHDASFADLSEEQ 127

Query: 83  MRQLLDYLLSKWQLMEESGHYRSVVLYRNFGPLSGGSLRHPHSQI 127
             R +LD    +   +        V  + N G   G +L HPH QI
Sbjct: 128 ARLVLDAWTDRTSELSHLPSVEQVFCFENRGAEIGVTLGHPHGQI 172
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1600

A DNA sequence (GBSx1694) was identified in *S.agalactiae* <SEQ ID 4943> which encodes the amino acid sequence <SEQ ID 4944>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10133> which encodes amino acid sequence <SEQ ID 10134> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06998 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 186/410 (45%), Positives = 258/410 (62%), Gaps = 27/410 (6%)

Query: 4   YDTIIIGGGPAGMMAAISSNFYGNKTLLIEKNKRLGKKLAGTGGGRCNVTNNGNLDELLA 63
            ++ I+IGGGPAG+MA++S+  +G + LL++K  +LG+KLA +GGGRCNVTN   LDEL+A
Sbjct: 2   HEVIVIGGGPAGLMASVSAAEHGARVLLLDKGDKLGRKLAISGGGRCNVTNRMPLDELIA 61

Query: 64  GIPGNGRFLYSVFSQFDNHDIINFFQDNGVTLKEEDHGRMFPTTDKSRTIINALENKIKE 123
             IPGNGRF+YS FS F+N DII FF+  G+ LKEED GRMFP +DK+ T++  L  +I +
Sbjct: 62  HIPGNGRFMYSPFSVFNNEDIIRFFERLGIALKEEDRGRMFPVSDKATTVVQTLLKRIND 121

Query: 124 LGGQIMTDTEVVSVK-KIGDSFYIKTKDTQFASDK-LIVTTGGKSYPSTGSTGFGHDIAR 181
            LG  + T+T V S++   G    ++ K+ +    K +IV TGG+S P TGSTG  +  A+
Sbjct: 122 LGVTVRTNTAVASLEYDDGRIAMVQLKNGERLKTKTVIVATGGQSVPHTGSTGDAYPWAK 181

Query: 182 HFKLEVTDMEAAESPLLTDFP---HKKLQGISLDDVTLSF----EKHIITH--DLLFTHF 232
                 +T++   E P+ +  P    KKLQG+SL D+ LS      K I TH  D++FTHF
Sbjct: 182 AAGHTITELYPTEVPITSAEPFIQEKKLQGLSLRDIELSVYAPNGKQIKTHDGDMIFTHF 241

Query: 233 GLSGPAALRISSFVKGGETIY--------LDVLPNISVKEL-EIHFQN---EREKSLKNA 280
            GLSGPAALR S +V      Y        +D+ P I  + L +   QN   E +K+LK
Sbjct: 242 GLSGPAALRCSQYVVKALKKYKQPTIEMRIDLRPTIPAEALFQETIQNIKAEPKKALKTV 301

Query: 281 LKILLPERLAEFYAEDL--PEKVKQVSVKD--LEMLIQKLKKLPILVTGKMSLAKSFVTK 336
            L+ + PER  ++  E L     +   SV+   +  ++Q+LK     V G +S+ K+FVT
Sbjct: 302 LRGIAPERFLQYIYERLRIDSNLPCASVRHEVIREIVQQLKSFSFHVNGTLSIEKAFVTG 361

Query: 337 GGVDLKEINPKTLESKKVAGLHFAGEVLDINAHTGGFNITSALCTGWVAG 386
            GGV +KEI PKT+ SKK AGL F GEVLDI+ +TGG+NIT A  TG+ AG
Sbjct: 362 GGVSVKEIEPKTMHSKKKAGLFFCGEVLDIHGYTGGYNITCAFSTGYTAG 411
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4945> which encodes the amino acid sequence <SEQ ID 4946>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0448(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 308/386 (79%), Positives = 344/386 (88%)

Query: 1   MKHYDTIIIGGGPAGMMAAISSNFYGNKTLLIEKNKRLGKKLAGTGGGRCNVTNNGNLDE 60
            M  YDTIIIGGGPAGMMAAISS++YG KTLLIEKN+RLGKKLAGTGGGRCNVTN+GNLD
Sbjct: 1   MTQYDTIIIGGGPAGMMAAISSSYYGYKTLLIEKNRRLGKKLAGTGGGRCNVTNSGNLDV 60

Query: 61  LLAGIPGNGRFLYSVFSQFDNHDIINFFQDNGVTLKEEDHGRMFPTTDKSRTIINALENK 120
            L+AGIPGNGRFLYSVFSQFDNHDII FF++NGV LKEEDHGRMFPTTDKSRTII+ALE K
Sbjct: 61  LMAGIPGNGRFLYSVFSQFDNHDIIAFFEENGVKLKEEDHGRMFPTTDKSRTIIDALEKK 120

Query: 121 IKELGGQIMTDTEVVSVKKIGDSFYIKTKDTQFASDKLIVTTGGKSYPSTGSTGFGHDIA 180
            IK LGGQ++T TEVVSVKK  D FY+K+ D  F   KLIVTTGGKSYPSTGSTGFGHDIA
Sbjct: 121 IKALGGQVLTSTEVVSVKKQDDLFYLKSADQTFTCQKLIVTTGGKSYPSTGSTGFGHDIA 180

Query: 181 RHFKLEVTDMEAAESPLLTDFPHKKLQGISLDDVTLSFEKHIITHDLLFTHFGLSGPAAL 240
            RHFKL VTD+EAAESPLLTDFPHK LQGISLDDVTLS++KH+ITHDLLFTHFGLSGPAAL
Sbjct: 181 RHFKLTVTDLEAAESPLLTDFPHKVLQGISLDDVTLSYDKHVITHDLLFTHFGLSGPAAL 240

Query: 241 RISSFVKGGETIYLDVLPNISVKELEIHFQNEREKSLKNALKILLPERLAEFYAEDLPEK 300
            R+SSFVKGGE   LD LP++S  +L  +  ++R+K++KNALK LLPER+A+F +ED PEK
Sbjct: 241 RLSSFVKGGEIAELDFLPHLSTDDLTAYLSDQRDKNIKNALKGLLPERVADFLSEDYPEK 300

Query: 301 VKQVSVKDLEMLIQKLKKLPILVTGKMSLAKSFVTKGGVDLKEINPKTLESKKVAGLHFA 360
```

```
           VKQ+S K  + L+ KLK L I +TGKMSLAKSFVTKGGVDLKEINPKTLESKKV GL+FA
Sbjct: 301 VKQLSPKQEKELLDKLKHLQIPITGKMSLAKSFVTKGGVDLKEINPKTLESKKVPGLYFA 360

Query: 361 GEVLDINAHTGGFNITSALCTGWVAG 386
           GEVLDINAHTGGFNITSALC+GW+AG
Sbjct: 361 GEVLDINAHTGGFNITSALCSGWIAG 386
```

SEQ ID 4944 (GBS196) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 3; MW 44.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 4; MW 69.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1601

A DNA sequence (GBSx1695) was identified in *S.agalactiae* <SEQ ID 4947> which encodes the amino acid sequence <SEQ ID 4948>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1550(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10131> which encodes amino acid sequence <SEQ ID 10132> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA73267 GB:Y12736 orfX [Lactococcus lactis subsp. cremoris]
 Identities = 51/173 (29%), Positives = 87/173 (49%), Gaps = 20/173 (11%)

Query: 19  KTVSELAEILGVSRQAMNNRV-KTLPEECVEK---NSKGVTVVNRDGLIKLEEIYKKTIL 74
           KT+ ELA+ LGVS+Q + N++ K   E+ V+          V+N  G      + KKT+
Sbjct: 6   KTIKELADELGVSKQTIRNKIDKDFREKFVQTIKIKGNNTLVINNAGY----SLLKKTLQ 61

Query: 75  EEEPIDEEASRRELLEILVDEKNTEITRLYEQLKAKDIQIASKDEQLHVKDIQIAEKDKQ 134
            +     +  + +  +  +      I  L EQL  K+ Q++ KD+QL  KD QI++
Sbjct: 62  NDTAQTAKTLQNDTAQTKL------ICFLEEQLDKKEQQLSVKDKQLENKDTQISQMQNL 115

Query: 135 LDQQQQLTLTAMEDTQRLQLELNEAKA------EVEEIQEAKEEKIQELEAVK 181
           LDQQQ+L L   +  +  + E+NE KA      ++++    + E  +E+E +K
Sbjct: 116 LDQQQRLALQDKKLLEEYKSEINELKALKMPREDMKDGSSIRGEAQEEIERLK 168
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4949> which encodes the amino acid sequence <SEQ ID 4950>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3951(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 132/194 (68%), Positives = 154/194 (79%), Gaps = 4/194 (2%)
```

```
Query: 1   MIFFYKKI---STKEEVMTVEKTVSELAEILGVSRQAMNNRVKTLPEECVEKNSKGVTVV 57
           M+ F  +I   S KEE M +EKTVSELA+ILGVSRQA+NNRVK+LPEE ++KN KGVTVV
Sbjct: 1   MVLFLIRIFSDSDKEENMGIEKTVSELADILGVSRQAVNNRVKSLPEEDLDKNEKGVTVV 60

Query: 58  NRDGLIKLEEIYKKTILEEEPIDEEASRRELLEILVDEKNTEITRLYEQLKAKDIQIASK 117
            R GL+KLEEIYKKTI ++EPI EE  +RELLEILVDEKNTEITRLYEQLKAKD Q+ASK
Sbjct: 61  KRSGLVKLEEIYKKTIFDDEPISEETKQRELLEILVDEKNTEITRLYEQLKAKDAQLASK 120

Query: 118 DEQLHVKDIQIAEKDKQLDQQQQLTLTAMEDTQRLQLELNEAKAEVEEIQEAKEEKIQEL 177
           DEQ+ VKD+QIAEKDKQLDQQQQLT  AM D + L+LEL EAKAE  + +  + E++Q
Sbjct: 121 DEQMRVKDVQIAEKDKQLDQQQQLTAKAMADKETLKLELEEAKAEANQAR-LQVEEVQAE 179

Query: 178 EAVKKSFFGRFFNK 191
              KK FF R F K
Sbjct: 180 VGPKKGFFTRLFAK 193
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1602

A DNA sequence (GBSx1697) was identified in *S.agalactiae* <SEQ ID 4951> which encodes the amino acid sequence <SEQ ID 4952>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2157(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06137 GB:AP001515 DNA polymerase III (alpha subunit)
           [Bacillus halodurans]
 Identities = 31/87 (35%), Positives = 52/87 (59%), Gaps = 1/87 (1%)

Query: 13  EYIAFDLEFNTVGE-HSHIIQVSAVKYSNHQEIALFDTYVHTKVPLQSFINGLTGITARD 71
           E++ FD+E   +   ++ II+++AVK  N + I  F+ +     PL + I  LTGIT
Sbjct: 418 EFVVFDVETTGLSAVYNKIIELAAVKVKNGEIIDRFERFADPHEPLTNTIIELTGITDDM 477

Query: 72  IIGAPKIEIVLTDFQSFVGDTPLIGYN 98
           + G P++E VL +F +F+GD  L+ +N
Sbjct: 478 LKGQPEVEQVLNEFHAFIGDAVLVAHN 504
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4953> which encodes the amino acid sequence <SEQ ID 4954>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3427(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 136/200 (68%), Positives = 159/200 (79%)

Query: 3   FLGEIMKQLQEYIAFDLEFNTVGEHSHIIQVSAVKYSNHQEIALFDTYVHTKVPLQSFIN 62
           FL E MK L  YIAFDLEFNTV + SHIIQVSAVKY +H+E+  FDTYV+T VPLQSFIN
Sbjct: 9   FLEENMKHLDTYIAFDLEFNTVNDVSHIIQVSAVKYDHHKEVDSFDTYVYTDVPLQSFIN 68
```

```
Query: 63  GLTGITARDIIGAPKIEIVLTDFQSFVGDTPLIGYNGYKSDLPLLVENGLDLTSQYQVDL 122
           GLTGIT+  I   PK+E V+   F++FVG+ PLIGYN  KSDLP+L ENGLDL  QYQ+DL
Sbjct: 69  GLTGITSDKIAAEPKVEEVMAAFKNFVGELPLIGYNAQKSDLPILAENGLDLRDQYQIDL 128

Query: 123 YDEAFVRRSTDLNGIVNLKLTTVADFLGIKGKAHNSLEDARMTARVYEKFLDLDENKIYL 182
           +DEA+ RRS DLNGI NL+L TVA FLGIKG+ HNSLEDARMTA +Y+ FL+ D NK YL
Sbjct: 129 FDEAYDRRSADLNGIANLRLQTVATFLGIKGRGHNSLEDARMTAVIYKSFLETDTNKAYL 188

Query: 183 KQQKEVAVDSPFATLGNLFD 202
            QQ+EV  D+PFA LG+ FD
Sbjct: 189 SQQEEVTTDNPFAALGDFFD 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1603

A DNA sequence (GBSx1698) was identified in *S.agalactiae* <SEQ ID 4955> which encodes the amino acid sequence <SEQ ID 4956>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.10    Transmembrane  143 - 159 ( 136 - 166)
   INTEGRAL    Likelihood = -4.73    Transmembrane  169 - 185 ( 168 - 188)

----- Final Results -----
               bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB42766 GB:AL049841 transcriptional regulator [Streptomyces
           coelicolor A3(2)]
 Identities = 46/141 (32%), Positives = 71/141 (49%), Gaps = 11/141 (7%)

Query: 5   YSTGDLAKEAGVTVRTVQYYDKRGILSPSELSEGGRRVYSIADLEKLRQIIYLRDLDFSI 64
           YS G +A  AGVTVRT+ +YD  G+L PSE S  G R YS ADL++L+QI++ R+L F +
Sbjct: 3   YSVGQVAGFAGVTVRTLHHYDDIGLLVPSERSHAGHRRYSDADLDRLQQILFYRELGFPL 62

Query: 65  DNIKNLFTEDNASQILELFLQVQIRELRL--------AIDSKKDKLDKAVNLLKTVEKQD 116
           D +  L  +  A     L  Q ++   R+        A++   +     +NL     ++
Sbjct: 63  DEVAALLDDPAADPRAHLRRQHELLSARIGKLQKMAAAVEQAMEARSMGINL---TPEEK 119

Query: 117 SKTLGYLSDIVLMEENKRKWG 137
            +  G        EE + +WG
Sbjct: 120 FEVFGDFDPDQYEEEVRERWG 140
```

There is also homology to SEQ ID 1712.

SEQ ID 4956 (GBS372) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 8; MW 55kDa).

GBS372-GST was purified as shown in Figure 215, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1604

A DNA sequence (GBSx1699) was identified in *S.agalactiae* <SEQ ID 4957> which encodes the amino acid sequence <SEQ ID 4958>. This protein is predicted to be cyclopropane-fatty-acyl-phospholipid synthase (mma2). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3145(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD07482 GB:AE000557 cyclopropane fatty acid synthase (cfa)
           [Helicobacter pylori 26695]
 Identities = 167/397 (42%), Positives = 254/397 (63%), Gaps = 14/397 (3%)

Query: 2   VMDSLIIKQLIKSTFDIPLQVTYPNGNIETYNGSNPHVKLKLNKNFSVSELSKDPSIVLG 61
           ++   ++K + K   +   QV + + ++      +P   LK+++    S++ KD S+ +
Sbjct: 1   MISKFLLKSMFKQWKNGDYQVVFWDNSVYRNGEHSPKFTLKIHRPLKFSDIKKDMSLTIA 60

Query: 62  EAVMDGDIEIYGSIQELILSAY-RCGDSFLRNSKFSKLIPKQFHDKKHSKSDIQKHYDIG 120
           EA MDG I+I GS+ E++ S Y +     L     +K I K   +     S+I KHYD+G
Sbjct: 61  EAYMDGVIDIEGSMDEVMHSLYLQTNYEHLHKHDNAKAIQKPIKES----SNISKHYDLG 116

Query: 121 NDFYKLWLDDTMTYSCAYFKHENDSLEQAQLNKVHHILNKLNAQPGGKLLDIGCGWGTLI 180
           NDFY +WLD+T++YSCAYFK ++D+L  AQL K+ H L KL+ +PG KLLDIGCGWG L
Sbjct: 117 NDFYSIWLDETLSYSCAYFKKDDDTLHAAQLQKLDHTLKKLHLKPGEKLLDIGCGWGYLS 176

Query: 181 ITAAKEYGLNATGITLSEEQASFITKRIKEEGLENKVTVLIKDYRDI---RETYDYITSV 237
           + AA+EYG    GIT+S EQ     KR++E GLE+KVT+ + +Y+D+      +D + SV
Sbjct: 177 VKAAQEYGAEVMGITISSEQYKQANKRVQELGLEDKVTIKLLNYQDLDGRLYRFDKVVSV 236

Query: 238 GMFEHVGKENLSQYFQTISKRLNINGLALIHGITGQVGGNHGSGTNSWINKYIFPGGYIP 297
           GMFEHVGK+NL  YF+ + + L   G+ L+H I     G     TN+W++KYIFPGGY+P
Sbjct: 237 GMFEHVGKDNLPFYFKKVKEVLKRGGMFLLHSILCCFEGK----TNAWVDKYIFPGGYLP 292

Query: 298 RLTENLNHIASAGLQIADLEPLRRHYQKTLELWTKNFHNALPEVQK-THDKRFINMWDLY 356
            L E ++ ++     +   E LR HY KTL++W  NF++ L +V++ ++D+RFI MWDLY
Sbjct: 293 SLREVMSVMSECDFHLLMAESLRIHYAKTLDIWRNNFNHNLDQVKRLSYDERFIRMWDLY 352

Query: 357 LQSCAASFESGNIDIFQYLLSKGVSKDTMPMTRDYMY 393
           L++CA++F  G+ D+FQ LL+  V  +T P+T++Y+Y
Sbjct: 353 LRTCASAFRVGSADLFQLLLTNSVD-NTFPLTKEYIY 388
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1605

A DNA sequence (GBSx1700) was identified in *S.agalactiae* <SEQ ID 4959> which encodes the amino acid sequence <SEQ ID 4960>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4903(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11796 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 44/97 (45%), Positives = 60/97 (61%)

Query: 1   MMNMQNMMRQAQKLQKQMEQKQADLAASQFTGKSAQELVTVTFTGDKKLISIDYKEAVVD 60
            M NMQ MM+Q QK+QK M + Q +LA     G +   +VTV   G K+++ +  KE VVD
Sbjct: 5   MGNMQKMMKQMQKMQKDMAKAQEELAEKVVEGTAGGGMVTVKANGQKEILDVIIKEEVVD 64

Query: 61  PEDIETLQDMTTQAINDALSQVDDATKKIMGAFAGKM 97
            PEDI+ LQD+    A N+AL +VD+ T + MG F   M
Sbjct: 65  PEDIDMLQDLVLAATNEALKKVDEITNETMGQFTKGM 101
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4961> which encodes the amino acid sequence <SEQ ID 4962>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4451(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 84/99 (84%), Positives = 94/99 (94%)

Query: 1  MMNMQNMMRQAQKLQKQMEQKQADLAASQFTGKSAQELVTVTFTGDKKLISIDYKEAVVD 60
           MMNMQNMM+QAQKLQKQMEQKQADLAA QFTGKSAQ+LVT TFTGDKKL+ ID+KEAVVD
Sbjct: 1  MMNMQNMMKQAQKLQKQMEQKQADLAAMQFTGKSAQDLVTATFTGDKKLVGIDFKEAVVD 60

Query: 61 PEDIETLQDMTTQAINDALSQVDDATKKIMGAFAGKMPF 99
           PED+ETLQDMTTQAINDAL+Q+D+ TKK +GAFAGK+PF
Sbjct: 61 PEDVETLQDMTTQAINDALTQIDETTKKTLGAFAGKLPF 99
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1606

A DNA sequence (GBSx1701) was identified in *S.agalactiae* <SEQ ID 4963> which encodes the amino acid sequence <SEQ ID 4964>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3963(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1607

A DNA sequence (GBSx1702) was identified in *S.agalactiae* <SEQ ID 4965> which encodes the amino acid sequence <SEQ ID 4966>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.76   Transmembrane   21 -  37 (  19 -  39)

----- Final Results -----
            bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10129> which encodes amino acid sequence <SEQ ID 10130> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1608

A DNA sequence (GBSx1703) was identified in *S.agalactiae* <SEQ ID 4967> which encodes the amino acid sequence <SEQ ID 4968>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.1783(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1609

A DNA sequence (GBSx1704) was identified in *S.agalactiae* <SEQ ID 4969> which encodes the amino acid sequence <SEQ ID 4970>. This protein is predicted to be probable 1,4-dihydroxy-2-naphthoate octaprenyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.76   Transmembrane  239 - 255 ( 219 - 260)
   INTEGRAL    Likelihood = -8.33   Transmembrane  221 - 237 ( 219 - 238)
   INTEGRAL    Likelihood = -6.74   Transmembrane   91 - 107 (  89 - 113)
   INTEGRAL    Likelihood = -6.32   Transmembrane   39 -  55 (  35 -  59)
   INTEGRAL    Likelihood = -3.77   Transmembrane  111 - 127 ( 111 - 132)
   INTEGRAL    Likelihood = -2.97   Transmembrane  144 - 160 ( 143 - 161)
   INTEGRAL    Likelihood = -1.28   Transmembrane  275 - 291 ( 275 - 291)
   INTEGRAL    Likelihood = -0.59   Transmembrane  177 - 193 ( 177 - 193)
```

```
----- Final Results -----
                 bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15875 GB:Z99123 alternate gene name: ipa-6d~similar to
           quinone biosynthesis [Bacillus subtilis]
 Identities = 75/290 (25%), Positives = 139/290 (47%), Gaps = 15/290 (5%)

Query: 5   IFLELVEMKAKTASVLPFLIGLCFSAYYYNSVHPVYVGLFFVAMFLFNMFVDIWNNYNDY 64
           I  +L      TAS +P L+G   + +Y      +++ + F  +++  +  +++N Y D+
Sbjct: 21  ILWQLTRPHTLTASFVPVLLGTVLAMFYVKVDLLLFLAMLFSCLWI-QIATNLFNEYYDF 79

Query: 65  RNAVDL-DYKNDTNIIGRENLSLRQIEVIMASLVITSSMIGLVLVSQVGLPLLWMGLFCF 123
           +  +D  +       I R  +  + I  +  +    + ++G+ + +     L  +GL
Sbjct: 80  KRGLDTAESVGIGGAIVRHGMKPKTILQLALASYGIAILLGVYICASSSWWLALIGLVGM 139

Query: 124 GIGVLYSFGPRPLSSLPLGEVFSGLTMGFMISLICVYLNTYQNFSWDILNLSKIFLISLP 183
            IG LY+ GP P++  P GE+FSG+ MG +  LI  ++ T      D +N+  I LIS+P
Sbjct: 140 AIGYLYTGGPLPIAYTPFGELFSGICMGSVFVLISFFIQT------DKINMQSI-LISIP 192

Query: 184 NTLWIANLMLANNLCDKEEDEKNHRYTLVHYTGIRGGLLLFAISNSIALLAIVFEFLFGL 243
             + +  + L+NN+ D EED+K  R TL    G +G + L A S ++A + +V   + G
Sbjct: 193 IAILVGAINLSNNIRDIEEDKKGGRKTLAILMGHKGAVTLLAASFAVAYIWVVGLVITGA 252

Query: 244 APVTVLLSLLLIPFIYKQTKLLWQKQVKRETFVCAVRILALGSATQVLTY 293
           A   + +  L +P   + K   Q ++          I+A+ S  Q  T+
Sbjct: 253 ASPWLFVVFLSVPKPVQAVKGFVQNEMPMN------MIVAMKSTAQTNTF 296
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1610

A DNA sequence (GBSx1705) was identified in *S.agalactiae* <SEQ ID 4971> which encodes the amino acid sequence <SEQ ID 4972>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane  155 - 171 ( 154 - 171)

----- Final Results -----
                 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15200 GB:Z99120 similar to NADH dehydrogenase [Bacillus subtilis]
 Identities = 178/403 (44%), Positives = 249/403 (61%), Gaps = 7/403 (1%)

Query: 3   EILVLGAGYAGLKAVRNLQKQSG--DFHITLVDMNDYHYEATELHEVAAGSQPKEKITFP 60
           +I++LGAGY GL  V  L K  G  D  ITLV+ ++YHYE T +HE +AG+   ++  +
Sbjct: 7   KIVILGAGYGGLMTVTRLTKYVGPNDADITLVNKHNYHYETTWMHEASAGTLHHDRCRYQ 66

Query: 61  IKDVINTNKVNFMQDEVLRVDAENKTVTVKNNGELHYDYVVVALGFVSETFGIKGAMENA 120
           IKDVIN ++VNF+QD V  +  + K V + N GEL YDY+V+ LG V ETFGIKG  E A
Sbjct: 67  IKDVINQSRVNFVQDTVKAIKIDEKKVVLAN-GELQYDYLVIGLGAVPETFGIKGLKEYA 125

Query: 121 LQMTNISQAENIHNHIVNTMKLYRETKDE--NLLKLLVCGAGFTGIELAGAMVDERPKYA 178
             + NI+ +  +  HI      Y    ++  + L ++V GAGFTGIE  G +    P+
```

```
Sbjct: 126 FPIANINTSRLLREHIELQFATYNTEAEKRPDRLTIVVGGAGFTGIEFLGELAARVPELC 185

Query: 179 ALAGVKPEQIEIICVEAATRILPMFDDELAQYGVNLIKDLGINLMLGSMIKEIKPGEVVY 238
               V    + IICVEAA  +LP FD EL  Y V+ +++ G+   +G+ ++E  P  V
Sbjct: 186 KEYDVDRSLVRIICVEAAPTVLPGFDPELVDYAVHYLEENGVEFKIGTAVQECTPEGVRV 245

Query: 239 GTSKEDEELKSITAGTIIWTTGVSGSPVMGESGFDQRRGRVMVNSDLRDPKYDNVYVIGD 298
           G  K+DEE + I + T++W  GV G P++ E+GF+  RGRV VN DLR P +DNV+++GD
Sbjct: 246 G--KKDEEPEQIKSQTVVWAAGVRGHPIVEEAGFENMRGRVKVNPDLRAPGHDNVFILGD 303

Query: 299 VSAFMDTESGRPFPTTAQIATRMGAHVAKNLLHQIKGEATEDFSYSPQGTVASVGNTHGL 358
            S FM+ ++ RP+P TAQIA + G  VAKNL   IKG   E+F    +GTVAS+G  + +
Sbjct: 304 SSLFMNEDTERPYPPTAQIAMQQGITVAKNLGRLIKGGELEEFKPDIKGTVASLGEHNAV 363

Query: 359 GVVGKTKIKKYPASVMKKIIMNKSLVDMGGLKELLAKGRFDLY 401
           GVV   K+K  PAS MKK+I N+SL  +GGL   L KG+F  +
Sbjct: 364 GVVYGRKLKGTPASFMKKVIDNRSLFMIGGLGLTLKKGKFKFF 406
```

There is also homology to SEQ ID 4666.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1611

A DNA sequence (GBSx1706) was identified in *S.agalactiae* <SEQ ID 4973> which encodes the amino acid sequence <SEQ ID 4974>. This protein is predicted to be cytochrome d ubiquinol oxidase, subunit I (cydA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.64    Transmembrane   19 -  35 (  15 -  38)
   INTEGRAL    Likelihood = -5.73    Transmembrane  226 - 242 ( 222 - 244)
   INTEGRAL    Likelihood = -4.94    Transmembrane  130 - 146 ( 126 - 149)
   INTEGRAL    Likelihood = -4.83    Transmembrane  429 - 445 ( 422 - 446)
   INTEGRAL    Likelihood = -3.77    Transmembrane   55 -  71 (  53 -  74)
   INTEGRAL    Likelihood = -3.56    Transmembrane  342 - 358 ( 340 - 359)
   INTEGRAL    Likelihood = -1.06    Transmembrane   89 - 105 (  89 - 106)
   INTEGRAL    Likelihood = -0.59    Transmembrane  186 - 202 ( 186 - 202)

----- Final Results -----
            bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15902 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit I)
          [Bacillus subtilis]
 Identities = 246/470 (52%), Positives = 319/470 (67%), Gaps = 12/470 (2%)

Query: 6   LARFQFAMTTVFHFFFVPFTIGTCLVVAIMETMYVITKNEEYKKLTKFWGNIMLLSFAVG 65
           LAR QFA TT+FHF FVP +IG   +VA+MET+Y++ KNE Y K+ KFWG++ L++FAVG
Sbjct: 6   LARIQFASTTLFHFLFVPMSIGLVFMVALMETLYLVKKNELYLKMAKFWGHLFLINFAVG 65

Query: 66  VVTGIIQEFQFGMNWSDYSRFVGDIFGAPLAIEALLAFFMESTFLGLWMFTWDNKKISKK 125
           VVTGI+QEFQFG+NWSDYSRFVGD+FGAPLAIEALLAFFMES F+GLW+F WD  ++ KK
Sbjct: 66  VVTGILQEFQFGLNWSDYSRFVGDVFGAPLAIEALLAFFMESIFIGLWIFGWD--RLPKK 123

Query: 126 LHVTFIWLVVFGSLMSAMWILTANSFMQHPVGYEVVNGRAQMTDFLALVKNPQFFYEFTH 185
           +H   IWLV FG++MS+ WILTANSFMQ PVG+ + NGRA+M DF AL+ NPQ + EF H
Sbjct: 124 IHALCIWLVSFGTIMSSFWILTANSFMQEPVGFTIKNGRAEMNDFGALITNPQLWVEFPH 183

Query: 186 VIFGAITMGGTVVAGMSAFRLLKSEQLKDTTVELYKKSVRIGLVVALLGSISVMGVGDLQ 245
           VIFGA+  G   +AG+SAF+LLK ++     V  +K+S ++ ++V L   + V   G +Q
```

```
Sbjct: 184 VIFGALATGAFFIAGVSAFKLLKKKE-----VPFFKQSFKLAMIVGLCAGLGVGLSGHMQ 238

Query: 246 MKALIHDQPMKFAAMEGDYEDSGDPAAWSVVAWANEAEHKQVFGIKIPYMLSILSYGKPS 305
            + L+  QPMK AA EG +EDSGDPAAW+  A  +    K    IK+PY LS L+Y K S
Sbjct: 239 AEHLMESQPMKMAASEGLWEDSGDPAAWTAFATIDTKNEKSSNEIKVPYALSYLAYQKFS 298

Query: 306 GSVKGMDTANKELVAKYGKDNYYPMVNLLFYGFRTMAAMGTAIMGVSVLGLFLTRKKKPI 365
           GSVKGM T   E    YGK +Y P V   F+ FR M   G  ++  ++ GL+L R+KK
Sbjct: 299 GSVKGMKTLQAEYEKIYGKGDYIPPVKTTFWSFRIMVGAGVVMILAALGGLWLNRRKK-- 356

Query: 366 LYKHKWMLWIVALTTFAPFLANTFGWIVTEQGRYPWTVYGLFKIKDSVSPNVSVASLFVS 425
           L   KW L I+      PFLAN+ GWI+TE GR PWTV GL     SVSPNV+  SL  S
Sbjct: 357 LENSKWYLRIMIALISFPFLANSAGWIMTEIGRQPWTVMGLMTTAQSVSPNVTAGSLLFS 416

Query: 426 NTVYFLLFGGLAVMMISLTIRELKKGPEYEDEHGHHGAYTSIDPFEEGAY 475
              + +++  L  +++ L IRE+KKG E+++   HH    S DPF +  Y
Sbjct: 417 IIAFGVMYMILGALLVFLFIREIKKGAEHDN---HHDVPVSTDPFSQEVY 463
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1612

A DNA sequence (GBSx1707) was identified in *S.agalactiae* <SEQ ID 4975> which encodes the amino acid sequence <SEQ ID 4976>. This protein is predicted to be cytochrome oxidase subunit II (cydB-1). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.49   Transmembrane  226 - 242 ( 220 - 250)
   INTEGRAL    Likelihood = -8.12   Transmembrane  254 - 270 ( 250 - 282)
   INTEGRAL    Likelihood = -7.64   Transmembrane  198 - 214 ( 196 - 218)
   INTEGRAL    Likelihood = -6.95   Transmembrane   85 - 101 (  76 - 103)
   INTEGRAL    Likelihood = -6.74   Transmembrane    6 -  22 (   1 -  27)
   INTEGRAL    Likelihood = -6.16   Transmembrane  300 - 316 ( 298 - 322)
   INTEGRAL    Likelihood = -5.36   Transmembrane  119 - 135 ( 117 - 143)
   INTEGRAL    Likelihood = -4.04   Transmembrane  159 - 175 ( 155 - 178)

----- Final Results -----
               bacterial membrane --- Certainty=0.6795(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15901 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit II)
          [Bacillus subtilis]
 Identities = 158/331 (47%), Positives = 223/331 (66%), Gaps = 1/331 (0%)

Query: 1   MSALQFFWFFLIGLLFSGFFFLEGFDFGVGMAVQTLTHNEHEKDQVVETIGPVWDGNEVW 60
           M++L   WF L+ +LF GFFFLEGFDFGVGMA + L HNE E+  ++ TIGP WD NEVW
Sbjct: 1   MASLHDLWFILVAVLFVGFFFLEGFDFGVGMATRFLGHNELERRVLINTIGPFWDANEVW 60

Query: 61  LLTGGGAMFASFPYWYASLFSGYYLILLTILFGLIIRGVSFEFRHKVPAEK-KQFWNWTL 119
           LLTG GA+FA+FP WYA++ SGYY+  + +L  L+ RGV+FEFR KV   K  + W+W +
Sbjct: 61  LLTGAGAIFAAFPNWYATMLSGYYIPFVIVLLALMGRGVAFEFRGKVDHLKWVKVWDWVV 120

Query: 120 TIGSAIVPFFFGIMFISLIQGMPLDASGNLSAQFSDYFNIFSLVGGVAMVLLAYLHGLNY 179
             GS I PF  G++F +L +GMP+DA  N+ A  SDY N++S++GGV + LL + HGL +
Sbjct: 121 FFGSLIPPFVLGVLFTTLFRGMPIDADMNIHAHVSDYINVYSILGGVTVTLLCFQHGLMF 180

Query: 180 IALKTEGPIRERARNYAQLLYWVLYLGLALFAVLLYFKTDFFSNHPIVTTIMVLVIVVLA 239
           I L+T G ++ RAR  AQ +  V+++ +  FA L  ++TD F+    +T  + ++IV+
```

```
Sbjct: 181 ITLRTIGDLQNRARKMAQKIMGVVFVAVLAFAALSAYQTDMFTRRGEITIPLAVLIVICF 240

Query: 240 VLAHASTFKGAEMTAFLASGLSLVSVVVLLFQGLFPRVMISSISPKYDLLIQNASSTPYT 299
           +LA     K +    F  +G  L   V ++F  LFPRVM+SS+   YDL + NASS  Y+
Sbjct: 241 MLAAVFIRKKKDGWTFGMTGAGLALTVGMIFISLFPRVMVSSLHSAYDLTVANASSGDYS 300

Query: 300 LKVMSIVAITLVPFVLAYTAWAYYIFRKRIT 330
           LKVMSI A+TL+PFV+    W+YY+FRKR++
Sbjct: 301 LKVMSIAALTLLPFVIGSQIWSYYVFRKRVS 331
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1613

A DNA sequence (GBSx1708) was identified in *S.agalactiae* <SEQ ID 4977> which encodes the amino acid sequence <SEQ ID 4978>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1614

A DNA sequence (GBSx1709) was identified in *S.agalactiae* <SEQ ID 4979> which encodes the amino acid sequence <SEQ ID 4980>. This protein is predicted to be transport ATP-binding protein cydc (cydD). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.82   Transmembrane  158 - 174 ( 144 - 182)
   INTEGRAL    Likelihood = -6.48   Transmembrane   15 -  31 (  14 -  34)
   INTEGRAL    Likelihood = -5.31   Transmembrane  243 - 259 ( 238 - 266)
   INTEGRAL    Likelihood = -2.55   Transmembrane  136 - 152 ( 134 - 152)
   INTEGRAL    Likelihood = -0.48   Transmembrane  263 - 279 ( 263 - 279)

----- Final Results -----
               bacterial membrane --- Certainty=0.7729(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15900 GB:Z99123 ABC membrane transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 279/569 (49%), Positives = 401/569 (70%), Gaps = 6/569 (1%)

Query: 2   LDKAVMRLSGIHKLLGLLAGLDVLQAIFIIGQAYYLSLSITGLWEGQKLSSQTVYILLFM 61
           + K + R  G+ ++L L+  L ++Q   II QA +LS ++TGL+ G+ ++S     I  F+
Sbjct: 1   MGKDLFRYKGMKRILTLITCLTLIQTAAIIMQAEWLSEAVTGLFNGKGITSLLPVIGFFL 60
```

```
Query: 62  VSYLGRHVIDYIKNRKLDDFSTAQSSLLRRQLLDKLFDLGPKVVQEQGTGNVVTMALDGV 121
           ++++ RH +   + + +  ++     + LR+  LD+LF LGP+  +++GTG +VT+A++G+
Sbjct: 61  IAFIARHGMTVARQKIVYQYAARTGADLRKSFLDQLFRLGPRFAKKEGTGQMVTLAMEGI 120

Query: 122 SLVENYLRLVLNKMINMSIIPWIILAYIFYLDIESGAILLIVFPLIIIFMIILGYAAQAK 181
           S    YL L L KM++M+I+P  ++ Y+F+ D  S  IL+   P++IIFMI+LG  AQ K
Sbjct: 121 SQFRRYLELFLPKMVSMAIVPAAVVIYVFFQDRTSAIILVAAMPILIIFMILLGLVAQRK 180

Query: 182 ADKQYESYQVLSNHFLDSLRGIDTLKYFGLSKRYGKSIYQTSESFRKATMSTLKIGILST 241
           AD+Q++SYQ LSNHF+DSLRG++TL++ GLSK + K+I+  SE +RKATMSTL++  LS+
Sbjct: 181 ADRQWKSYQRLSNHFVDSLRGLETLRFLGLSKSHSKNIFYVSERYRKATMSTLRVAFLSS 240

Query: 242 FALDFFTTLSIAIVAVFLGLRLLNEQIYLLPALTILILSPEYFLPVRDFSSDYHATLDGK 301
           FALDFFT LS+A VAVFLGLRL++  I L PALT LIL+PEYFLPVR+  +DYHATL+G+
Sbjct: 241 FALDFFTMLSVATVAVFLGLRLIDGDILLGPALTALILAPEYFLPVREVGNDYHATLNGQ 300

Query: 302 NAFQAIQKVLNKTGIKGE-QLVIDDWSKESRLDLENIAIAYDQKRVVEDVTLRFRGHQKV 360
            A + IQ++L++ G K E  L ++ WS +  L L  +++     R V D+ L F+G +K+
Sbjct: 301 EAGKTIQEILSQPGFKEETPLQLEAWSDQDELKLSGVSVG----RSVSDIHLSFKGKKKI 356

Query: 361 ALVGVSGSGKSSLINLLSGFLGPDNGSLKVDGREVTNLDQEDWHKQMIYIPQTPYVFEMS 420
            ++G SG+GKS+LI++L GFL PD G ++V+G   ++L    W K ++YIPQ PY+F+ +
Sbjct: 357 GIIGASGAGKSTLIDILGGFLEPDGGMIEVNGTSRSHLQDGSWQKNLLYIPQHPYIFDDT 416

Query: 421 LRDNITFYTPNASDEEVVRAIHMVGLDSLLSELPDGLETRIGNGARPLSGGQAQRIALAR 480
           L +NI FY P+AS E+  RA    GL  L++ LPDGLE RIG G R LSGGQAQR+ALAR
Sbjct: 417 LGNNIRFYHPSASAEDTTRAAASAGLTELVNNLPDGLEGRIGEGGRALSGGQAQRVALAR 476

Query: 481 AFLDQNRRIMVFDEPTAHLDIETELELKEKMLPLMSDRLVIFATHRLHWLNQMDVIVVME 540
           AFL  NR I++ DEPTAHLDIETE E+KE ML L  D+LV  ATHRLHW+  MD I+V++
Sbjct: 477 AFLG-NRPILLLDEPTAHLDIETEYEIKETMLDLFEDKLVFLATHRLHWMLDMDEIIVLD 535

Query: 541 KGRVAEVGSYQELLAKKGYLYQLKHAMGG 569
            GRVAE+G++ ELL K G   +L  A  G
Sbjct: 536 GGRVAEIGTHNELLEKNGVYTKLVKAQLG 564
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4981> which encodes the amino acid sequence <SEQ ID 4982>. Analysis of this protein sequence reveals the following:

```
      Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.61   Transmembrane  159 - 175 ( 154 - 190)
   INTEGRAL    Likelihood =-10.03   Transmembrane   70 -  86 (  63 -  91)
   INTEGRAL    Likelihood = -3.03   Transmembrane  282 - 298 ( 282 - 301)
   INTEGRAL    Likelihood = -1.44   Transmembrane  261 - 277 ( 260 - 278)

----- Final Results -----
               bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC22320 GB:U32749 ATP-binding transport protein (cydD)
          [Haemophilus influenzae Rd]
 Identities = 167/544 (30%), Positives = 279/544 (50%), Gaps = 15/544 (2%)

Query: 46  MISFYLIAKTFSTFILGHAIALGRLAGLLLLLNVVGFVLAILGK---QLQGIASQFARDS 102
           + S+ L A  F    L  A+ LG +    L L       A  GK   Q    AS   +
Sbjct: 17  VFSYILQAAYFHELSLLSAVILGIVLIAALALR------AFAGKKSVQASYFASTKVKHE 70

Query: 103 LKQSFFEAFIDLDGQFDAHASDADILTLASQGIDSLDTYYGYYL-SLSMRTKWNCTTIMI 161
           L+   +     +        S + I+ +AS+G++ L+ Y+G YL  L        T
Sbjct: 71  LRSLIYRKLASMPLNQVNQQSTSSIIQVASEGVEQLEIYFGRYLPQLFYSLLAPLTLFAF 130

Query: 162 LVFLIYPLAGLVFLGVLPLIPLSIVAMQKRSQPNMSHYWSSYMDVGNLFMDDLKGLNTLY 221
```

```
           L+F  +  A ++ L  +PLIP+SI+A+ K ++  ++ YWS Y+ +G+ F+D+L+GL TL
Sbjct: 131 LIFFSFKTA-IILLICVPLIPMSIIAVNKIAKKLLAKYWSIYVGLGSSFLDNLQGLITLK 189

Query: 222 SYQATERYEQEFSGKAEQFRKATMSLLGFQLQAVGYMDAVMYLGIGLSGFLAVQALATGQ 281
            YQ      +    +AE FRK TM +L  QL +V  MD + Y G  +    A+      Q
Sbjct: 190 IYQDDAYKAKAMDKEAEHFRKITMKVLTMQLNSVSLMDLLAYGGAAIGILTALLQFQNAQ 249

Query: 282 LSFFNFLFFLLIATEFFTPIREQGYGMHLVMMNTKMADRIFSFLDS-VPARKDNKSKTAI 340
           LS    + F+L+++EFF P+R  G   H+ M     +D+IF+ LD+ V  ++      A
Sbjct: 250 LSVLGVILFILLSSEFFIPLRLLGSFFHVAMNGKAASDKIFTLLDTPVETQQSAVDFEAK 309

Query: 341 NFNQIDIQNISLAY-EKKTVLSGVTMTLTKGQLTAIAGVSGQGKTSLAQLLLKRQSATTG 399
           N  Q++I+++  +Y E+K  ++G+ +++   QL+   G SG GK++L  LL+    A  G
Sbjct: 310 NNVQVEIKDLHFSYSEEKPAITGLNLSILPNQLSVFVGKSGCGKSTLVSLLMGFNKAQQG 369

Query: 400 HILFDGLDSDNLSQETINQQVLYVSDQSTLLNRSIYDNLRLA-ANLSKKEILDWIDQHGL 458
            ILF+G ++ N+ + +  Q+V  VS  S +   ++ +N+ +A  + + ++I   ++Q  L
Sbjct: 370 EILFNGQNALNIDRTSFYQKVSLVSHSSYVFKGTLRENMTMAKIDATDEQIYACLEQVNL 429

Query: 459 LSFINWLPDGLDTIVGENGNLLSPGQKQQVICARALLSKRSLYIFDEATSSLDAENERII 518
             F+     GLD  +   G  LS GQ Q++  ARALL    LYIFDEATS++D E+E II
Sbjct: 430 AQFVR-DNGGLDMQLLSRGANLSGGQIQRLALARALLHNAELYIFDEATSNIDVESEEII 488

Query: 519 DNLITRLAKTAIVIVITHKMSRLKGANQVLFLNTGQPACLGKPCDLYRDQPTYRHLVDTQ 578
              I + +   +++I+H+++    A+ +  L+ G+    G   +L   Q  Y  +   Q
Sbjct: 489 LQFIQQFKQQKTIVMISHRLANAVNADCINVLDQGKLIEQGTHKELMEKQGAYAEMFQQQ 548

Query: 579 ARLE 582
             LE
Sbjct: 549 KDLE 552
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 143/552 (25%), Positives = 260/552 (46%), Gaps = 12/552 (2%)

Query: 1   MLDKAVMRLSGIHKLLGLLAGLDVLQAIFIIGQAYYLSLSITGLWEGQKLSSQTVYILLF 60
           +L +   R++    LL + A L  LQ +  +   Y ++ + +    G  ++   +  LL
Sbjct: 16  LLKRLRERIAPKRYLLYVSAFLSWLQFVMRMISFYLIAKTFSTFILGHAIALGRLAGLLL 75

Query: 61  MVSYLGRHVIDYIKNRKLDDFSTAQSSLLRRQLLDKLFDLGPKVVQEQGTGNVVTMALDG 120
           +++ +G  V+  +  +     S      L++   +   DL  +        +++T+A  G
Sbjct: 76  LLNVVG-FVLAILGKQLQGIASQFARDSLKQSFFEAFIDLDGQFDAHASDADILTLASQG 134

Query: 121 VSLVENYLRLVLNKMINMSIIPWIILAYIFYLDIESGAILLIVFPLIIIFMIILGYAAQA 180
           +  ++ Y    L+  +        I+  +F +   +G + L V PLI + ++ +   +Q
Sbjct: 135 IDSLDTYYGYYLSLSMRTKWNCTTIMILVFLIYPLAGLVFLGVLPLIPLSIVAMQKRSQP 194

Query: 181 KADKQYESYQVLSNHFLDSLRGIDTLKYFGLSKRYGKSIYQTSESFRKATMSTLKIGILS 240
                + SY  + N F+D L+G++TL  +  ++RY +     +E FRKATMS L   + +
Sbjct: 195 NMSHYWSSYMDVGNLFMDDLKGLNTLYSYQATERYEQEFSGKAEQFRKATMSLLGFQLQA 254

Query: 241 TFALDFFTTLSIAIVAVFLGLRLLNEQIYLLPALTILILSPEYFLPVRDFSSDYHATLDG 300
              +D    L I +        L   Q+     L  L+++ E+F P+R+     H  +
Sbjct: 255 VGYMDAVMYLGIGLSGFLAVQALATGQLSFFNFLFFLLIATEFFTPIREQGYGMHLVMMN 314

Query: 301 KNAFQAIQKVLNKTGIKGEQLVIDDWSKE----SRLDLENIAIAYDQKRVVEDVTLRFRG 356
                 I   L+    +      D+ SK     +++D++NI++AY++K V+  VT+
Sbjct: 315 TKMADRIFSFLDSVPARK-----DNKSKTAINFNQIDIQNISLAYEKKTVLSGVTMTLTK 369

Query: 357 HQKVALVGVSGSGKSSLINLLSGFLGPDNGSLKVDGREVTNLDQEDWHKQMIYIPQTPYV 416
            Q  A+ GVSG GK+SL  LL        G +  DG +  NL QE  ++Q++Y+     +
Sbjct: 370 GQLTAIAGVSGQGKTSLAQLLLKRQSATTGHILFDGLDSDNLSQETINQQVLYVSDQSTL 429

Query: 417 FEMSLRDNITFYTPNASDEEVVRAIHMVGLDSLLSELPDGLETRIGNGARPLSGGQAQRI 476
              S+ DN+     N S +E++  I   GL S ++ LPDGL+T +G     LS GQ Q++
Sbjct: 430 LNRSIYDNLRL-AANLSKKEILDWIDQHGLLSFINWLPDGLDTIVGENGNLLSPGQKQQV 488

Query: 477 ALARAFLDQNRRIMVFDEPTAHLDIETELELKEKMLPLMSDRLVIFATHRLHWLNQMDVI 536
```

```
             ARA L + R + +FDE T+ LD E E  +   +  L    +VI  TH++  L   + +
Sbjct: 489 ICARALLSK-RSLYIFDEATSSLDAENERIIDNLITRLAKTAIVIVITHKMSRLKGANQV 547

Query: 537 VVMEKGRVAEVG 548
           + +  G+ A +G
Sbjct: 548 LFLNTGQPACLG 559
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1615

A DNA sequence (GBSx1710) was identified in *S.agalactiae* <SEQ ID 4983> which encodes the amino acid sequence <SEQ ID 4984>. This protein is predicted to be transport ATP-binding protein cydd (cydC). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.84    Transmembrane  260 - 276 ( 258 - 284)
   INTEGRAL    Likelihood = -9.34    Transmembrane  172 - 188 ( 147 - 199)
   INTEGRAL    Likelihood = -6.53    Transmembrane  150 - 166 ( 147 - 171)
   INTEGRAL    Likelihood = -6.05    Transmembrane   31 -  47 (  29 -  52)
   INTEGRAL    Likelihood = -3.35    Transmembrane   68 -  84 (  67 -  84)
   INTEGRAL    Likelihood = -1.17    Transmembrane  293 - 309 ( 292 - 310)
   INTEGRAL    Likelihood = -0.69    Transmembrane  494 - 510 ( 493 - 510)

----- Final Results -----
               bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10127> which encodes amino acid sequence <SEQ ID 10128> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15899 GB:Z99123 ABC membrane transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 262/573 (45%), Positives = 389/573 (67%), Gaps = 14/573 (2%)

Query: 16  LKTDQWIKPFFKQYKVSLVIALFLGFMTFFSASALMFNSGYLISKSASLPSNILLVYVPI 75
           +K ++WI P+ KQ     V+ +FLG +T FSA+ LMF SG+LISK+A+ P NILL+YVPI
Sbjct: 1   MKKEEWILPYIKQNARLFVLVIFLGAVTIFSAAFLMFTSGFLISKAATRPENILLIYVPI 60

Query: 76  VLTRAFGIGRPVFRYIERLTSHNWVLRMTSQLRLKLYHSLESNAIFMKRDFRLGDVMGLL 135
           V  R FGI R V RY+ERL  H+ +L++ S +R++LY+ LE  A+ ++  FR GD++G+L
Sbjct: 61  VAVRTFGIARSVSRYVERLVGHHIILKIVSDMRVRLYNMLEPGALMLRSRFRTGDMLGIL 120

Query: 136 AEDINYLQNLYLRTIFPTIIAWILYSFIIIATGFFSLWFALMMLLYLAIMIFLFPLWSIL 195
           +EDI +LQ+ +L+TIFP I A +LY+  +IA GFFS  FA+++ LYL +++ LFP+ S+L
Sbjct: 121 SEDIEHLQDAFLKTIFPAISALLLYAVSVIALGFFSWPFAILLALYLFVLVVLFPVVSLL 180

Query: 196 ANGARQTREKELKNHLYTDLTDNVLGISDWIFSQRGQEYVALHERSESELMAVQKKIRSF 255
              A+  + K  +N LY+ LTD V+G+SDW+FS R   ++  +E+ E +   +++K + F
Sbjct: 181 VTRAKNAKLKSGRNVLYSRLTDAVMGVSDWMFSGRRHAFIDAYEKEERDWFELERKKQRF 240

Query: 256 DNRRALIVELVFGFLAILVIIWASNQFIGHRGGEA--NWIAAFVLTVFPLSEAFAGLSAA 313
              R    + +   L +L++ W + Q      GE     IAAFVL VFPL+EAF  LS A
Sbjct: 241 TRWRDFAAQCLVAGLILLMLFWTAGQ---QADGELAKTMIAAFVLVVFPLTEAFLPLSDA 297

Query: 314 AQETNKYSDSIHRLN------ELSETYFETTQNQLPNKPYDFSVKNLSFQYKPQEKWVLH 367
             E   Y DSI R+N      E S+T  E+    L  +    + ++++F Y    + VLH
Sbjct: 298 LGEVPGYQDSIRRMNNVAPQPEASQT--ESGDQILDLQDVTLAFRDVTFSYDNSSQ-VLH 354
```

```
Query: 368 HLDLDIKEGEKIAILGRSGSGKSTLASLLRGDLKASQGEITLGDADVSIVGDCISNYIGV 427
           +     +++GEK+A+LGRSGSGKST  +L+ G LK   G +TL   + +++ D I++ + V
Sbjct: 355 NFSFTLRQGEKMALLGRSGSGKSTSLALIEGALKPDSGSVTLNGVETALLKDQIADAVAV 414

Query: 428 IQQAPYLFNTTLLNNIRIGNQDASEEDVWKVLERVGLKEMVTDLSDGLYTMVDEAGLRFS 487
           + Q P+LF+T++LNNIR+GN +AS+EDV +   ++V L + +  L DG +T V E G+RFS
Sbjct: 415 LNQKPHLFDTSILNNIRLGNGEASDEDVRRAAKQVKLHDYIESLPDGYHTSVQETGIRFS 474

Query: 488 GGERHRIALARILLKDVPIVILDEPTVGLDPITEQALLRVFMKELEGKTLVWITHHLKGI 547
           GGER RIALARILL+D PI+ILDEPTVGLDPITE+ L+     + L+GKT++WITHHL G+
Sbjct: 475 GGERQRIALARILLQDTPIIILDEPTVGLDPITERELMETVFEVLKGKTILWITHHLAGV 534

Query: 548 EHADRILFIENGQLELEGSPQELSQSSQRYRQL 580
           E AD+I+F+ENG+ E+EG+ +EL  +++RYR+L
Sbjct: 535 EAADKIVFLENGKTEMEGTHEELLAANERYRRL 567
```

A related GBS gene <SEQ ID 8861> and protein <SEQ ID 8862> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:    -15.90
GvH: Signal Score (-7.5): 1.97
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value: -12.84 threshold:  0.0
   INTEGRAL    Likelihood =-12.84   Transmembrane  260 - 276 ( 258 - 284)
   INTEGRAL    Likelihood = -9.34   Transmembrane  172 - 188 ( 147 - 199)
   INTEGRAL    Likelihood = -6.53   Transmembrane  150 - 166 ( 147 - 171)
   INTEGRAL    Likelihood = -6.05   Transmembrane   31 -  47 (  29 -  52)
   INTEGRAL    Likelihood = -3.35   Transmembrane   68 -  84 (  67 -  84)
   INTEGRAL    Likelihood = -1.17   Transmembrane  293 - 309 ( 292 - 310)
   INTEGRAL    Likelihood = -0.69   Transmembrane  494 - 510 ( 493 - 510)
   PERIPHERAL  Likelihood =  3.29     412
 modified ALOM score:   3.07

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00997(346 - 2052 of 2364)
EGAD|98910|BS3866(1 - 571 of 575) transport ATP-binding protein cydd {Bacillus subtilis}
OMNI|NT01BS4517 ABC transporter CydC, putative SP|P94367|CYDD_BACSU TRANSPORT ATP-BINDING
PROTEIN CYDD. GP|1783253|dbj|BAA11730.1||D83026 homologous to many ATP-binding transport
proteins; hypothetical {Bacillus subtilis} GP|2636408|emb|CAB15899.1||Z99123 ABC membrane
transporter (ATP-binding protein) {Bacillus subtilis} PIR|D69611|D69611 ABC transporter
required for expression of cytochrome bd (ATP-) cydD - Bacillus subtilis
%Match = 31.9
%Identity = 45.2  %Similarity = 69.1
Matches = 257  Mismatches = 172  Conservative Sub.s = 136

300       330       360       390       420       450       480       510
LKKDISIN*SMLWEEMMFKIPLFKELKTDQWIKPFFKQYKVSLVIALFLGFMTFFSASALMFNSGYLISKSASLPSNILL
                         :| ::|| |: ||    :|: :||| :| |||: ||| ||:||||:|: | ||||
                         MKKEEWILPYIKQNARLFVLVIFLGAVTIFSAAFLMFTSGFLISKAATRPENILL
                                  10        20        30        40        50

540       570       600       630       660       690       720       750
VYVPIVLTRAFGIGRPVFRYIERLTSHNWVLRMTSQLRLKLYHSLESNAIFMKRDFRLGDVMGLLAEDINYLQNLYLRTI
:|||||  | ||| | | ||:|||  |: :|:: | :|::||: ||  |: ::  || ||::|:|:||| :||: :|:||
IYVPIVAVRTFGIARSVSRYVERLVGHHIILKIVSDMRVRLYNMLEPGALMLRSRFRTGDMLGILSEDIEHLQDAFLKTI
            70        80        90       100       110       120       130
```

```
780       810       840       870       900       930       960       990
FPTIIAWILYSFIIIATGFFSLWFALMMLLYLAIMIFLFPLWSILANGARQTREKELKNHLYTDLTDNVLGISDWIFSQR
|| | | :||:  :|| ||||  ||::: ||| ::: |||: |:|   |:  : |  :| ||: ||| |:|:|||:|| |
FPAISALLLYAVSVIALGFFSWPFAILLALYLFVLVVLFPVVSLLVTRAKNAKLKSGRNVLYSRLTDAVMGVSDWMFSGR
            150       160       170       180       190       200       210

1020      1050      1080      1110      1140                1194      1224
GQEYVALHERSESELMAVQKKIRSFDNRRALIVELVFGFLAILVIIWASNQFIGHRGGE--ANWIAAFVLTVFPLSEAFA
 : ::  :|: | :   :::| : |   | :  : :   | :|:: | : |   :  ||     |||||| ||||:|||
RHAFIDAYEKEERDWFELERKKQRFTRWRDFAAQCLVAGLILLMLFWTAGQ---QADGELAKTMIAAFVLVVFPLTEAFL
            230       240       250       260          270       280       290

1254      1284      1302      1332      1362      1392      1422      1452
GLSAAAQETNKYSDSIHRLNELS----ETYFETTQNQLPNKPYDFSVKNLSFQYKPQEKWVLHHLDLDIKEGEKIAILGR
 || |  |   | ||| |:| ::     :  |:    |  :   :: ::::| |      |||:: : :::|||:|:|||
PLSDALGEVPGYQDSIRRMNNVAPQPEASQTESGDQILDLQDVTLAFRDVTFSY-DNSSQVLHNFSFTLRQGEKMALLGR
               310       320       330       340        350       360       370

1482      1512      1542      1572      1602      1632      1662      1692
SGSGKSTLASLLRGDLKASQGEITLGDADVSIVGDCISNYIGVIQQAPYXFNTTLLNTFRIGNQDASEEDVWKVLERVGL
|||||||  :|: | ||   | :||   : ::: | |:: : |: | |: |:|::||  |:|| :||:|||| :  ::| |
SGSGKSTSLALIEGALKPDSGSVTLNGVETALLKDQIADAVAVLNQKPHLFDTSILNNIRLGNGEASDEDVRRAAKQVKL
                390       400       410       420       430       440       450

1722      1752      1782      1812      1842      1872      1902      1932
KEMVTDLSDGLYTMVDEAGLRFSGGERHRIALARILLKDVPIVILDEPTVGLDPITEQALLRVFMKELEGKTLVWITHHL
 : :  | || :| | | |:|||||||:|||||||||:| ||:||||||||||||||: |:    : |:|||::||||||
HDYIESLPDGYHTSVQETGIRFSGGERQRIALARILLQDTPIIILDEPTVGLDPITERELMETVFEVLKGKTILWITHHL
                470       480       490       500       510       520       530

1962      1992      2022      2052      2082      2112      2142      2172
KGIEHADRILFIENGQLELEGSPQELSQSSQRYRQLKASDDGDL**LIGAINK***KNIP*LLF*HCGMFFYYLNFAF*K
 |:| ||:|:|:|||: |:||: :||  :::|||:|   |
AGVEAADKIVFLENGKTEMEGTHEELLAANERYRRLYHLDVPVK
                550       560       570
```

There is also homology to SEQ ID 478.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1616

A DNA sequence (GBSx1711) was identified in *S.agalactiae* <SEQ ID 4987> which encodes the amino acid sequence <SEQ ID 4988>. This protein is predicted to be spore germination protein C3 (ispB). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06   Transmembrane  111 - 127 ( 111 - 128)

----- Final Results -----
               bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14190 GB:Z99115 heptaprenyl diphosphate synthase component II
          [Bacillus subtilis]
 Identities = 101/318 (31%), Positives = 184/318 (57%), Gaps = 5/318 (1%)

Query: 8   YPELKKNIDETNQLIQERIQVRNKDIEAALSQLTAAGGKQLRPAFFYLFSQLGNKENQDT 67
            Y  L  +ID   + +++ ++      +  A    L   AGGK++RP  F  L    G+    D
Sbjct: 35  YSFLNDDIDVIERELEQTVRSDYPLLSEAGLHLLQAGGKRIRPVFVLLSGMFGD---YDI 91
```

```
Query: 68  QQLKKIAASLEILHVATLIHDDVIDDSPLRRGNMTIQSKFGKDIAVYTGDLLFTVFFDLI 127
            ++K +A +LE++H+A+L+HDDVIDD+ LRRG  TI++K+   IA+YTGD +     +++
Sbjct: 92  NKIKYVAVTLEMIHMASLVHDDVIDDAELRRGKPTIKAKWDNRIAMYTGDYMLAGSLEMM 151

Query: 128 LESMADTPFMRINAKSMRKILMGELDQMHLRYNQQQGIHHYLRAISGKTAELFKLASKEG 187
               + +    RI ++++ ++ +GE++Q+  +YN +Q +  YLR I  KTA L  ++ + G
Sbjct: 152 TR-INEPKAHRILSQTIVEVCLGEIEQIKDKYNMEQNLRTYLRRIKRKTALLIAVSCQLG 210

Query: 188 AYFGGAEKEVVRLAGHIGFNIGMTFQILDDILDYTADKKTFNKPVLEDLAQGIYSLPLLL 247
           A   GA++++ +     G+ +GM++QI+DDILD+T+ ++   KPV  DL QG  +LP+L
Sbjct: 211 AIASGADEKIHKALYWFGYYVGMSYQIIDDILDFTSTEEELGKPVGGDLLQGNVTLPVLY 270

Query: 248 AIEENPDIFKPILDKKTDMATEDMEKIAYLVVSHRGVDKARHLARKFTEKAISDINKLPQ 307
           A+ +NP +   +    ++   E +E I   +     ++ +  ++  + +KA   +N LP+
Sbjct: 271 AL-KNPALKNQLKLINSETTQEQLEPIIEEIKKTDAIEASMAVSEMYLQKAFQKLNTLPR 329

Query: 308 SSAKKQLLQLTNYLLKRK 325
             A+  L  +  Y+ KRK
Sbjct: 330 GRARSSLAAIAKYIGKRK 347
```

There is also homology to SEQ ID 284. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 65/227 (28%), Positives = 98/227 (42%), Gaps = 9/227 (3%)

Query: 43  AGGKQLRPAFFYLFSQLGNKENQDTQQLKKIAASLEILHVATLIHDDV--IDDSPLRRGN 100
           +GGK++RP       +      Q+      +AA+LE++H  +LIHDD+  +D+   RRG
Sbjct: 36  SGGKRIRPLILLEMIEGFGVSLQNAHF--DLAAALEMIHTGSLIHDDLPAMDNDDYRRGR 93

Query: 101 MTIQSKFGKDIAVYTGDLLFTVFFDLILESM--ADTPFMRINAKSMRKILMGELDQMHLR 158
           +T   +FG+  A+  GD LF   F LI ++   ++     I   S+     G +    L
Sbjct: 94  LTNHKQFGEATAILAGDSLFLDPFGLIAQAELNSEVKVALIQELSLASGTFGMVGGQMLD 153

Query: 159 Y---NQQQGIHHYLRAISGKTAELFKLASKEGAYFGGAEKEVVRLAGHIGFNIGMTFQIL 215
               NQ   +         KT +L     K  A        V +     G  IG  FQI
Sbjct: 154 MKGENQALSLPQLSLIHLNKTGKLLAFPFKAAALITEQAMTVRQQLEQAGMLIGHAFQIR 213

Query: 216 DDILDYTADKKTFNKPVLEDLAQGIYSLPLLLAIEENPDIFKPILDK 262
           DDILD TA  +   K   +DL     + P LL +E +  +    LD+
Sbjct: 214 DDILDVTASFEDLGKTPKKDLFAEKATYPSLLGLEASYQLLTESLDQ 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1617

A DNA sequence (GBSx1712) was identified in *S.agalactiae* <SEQ ID 4989> which encodes the amino acid sequence <SEQ ID 4990>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3995(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25232 GB:M58315 dipeptidyl peptidase IV [Lactococcus lactis]
 Identities = 385/767 (50%), Positives = 504/767 (65%), Gaps = 21/767 (2%)

Query: 1    MRYNQFSYIPTKPNEAFEELKGLGFPLNKKNSDKANLEAFLRHSFLNQTDTDYALSLLIV 60
            MR+N FS +    +E   EL  LGF  +    +K  L+ FL  S  + TD       L
Sbjct: 1    MRFNHFSIVDKNFDEQLAELDQLGFRWSVFWDEKKILKDFLIQSPSDMTD-------LQA 53
```

```
Query: 61  DAKTDALTFFKSNSDLTLENLQWIYLQLLGFIPFVDFKDPKAF-------LQDINFPVSY 113
            A+ D + F KS+ +L  E    I LQLL F+P  DF+  KAF       L  I    ++
Sbjct: 54  TAELDVIEFLKSSIELDWEIFWNIALQLLDFVPNFDFEIGKAFEYAKNSNLPQIEAEMTT 113

Query: 114 DNIFQSLHHLLACRGKSGNTLIDQLVADGLLHADNHYHFFNGKSLATFNTNQLIREVVYV 173
           +NI  + ++LL  R K+G  L++  V++GLL  DNHYHFFN KSLATF+++ L REV++V
Sbjct: 114 ENIISAFYYLLCTRRKNGMILVEHWVSEGLLPLDNHYHFFNDKSLATFDSSLLEREVLWV 173

Query: 174 ETSLDTMSSGEHDLVKVNIIRPTTEHTIPTMMTASPYHQGINDPAADQKTYQMEGALAVK 233
           E+ +D+   GE+DL+K+ IIRP +   +P +MTASPYH GIND A D   + M   L  K
Sbjct: 174 ESPVDSEQRGENDLIKIQIIRPKSTEKLPVVMTASPYHLGINDKANDLALHDMNVELEEK 233

Query: 234 QPKHIQVDTKPFKEEVKHPSKLPI-SPATESFTHIDSYSLNDYFLSRGFANIYVSGVGTA 292
               I V+ K  ++      +LPI   A   FTH  +YSLNDYFL+RGFA+IYV+GVGT
Sbjct: 234 TSHEIHVEQKLPQKLSAKAKELPIVDKAPYRFTHGWTYSLNDYFLTRGFASIYVAGVGTR 293

Query: 293 GSTGFMTSGDYQQIQSFKAVIDWLNGKVTAFTSHKRDKQVKANWSNGLVATTGKSYLGTM 352
            S GF TSGDYQQI S  AVIDWLNG+  A+TS K+  ++KA+W+NG VA TGKSYLGTM
Sbjct: 294 SSDGFQTSGDYQQIYSMTAVIDWLNGRARAYTSRKKTHEIKASWANGKVAMTGKSYLGTM 353

Query: 353 STGLATTGVEGLKVIIAEAAISTWYDYYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAG 412
           + G ATTGVEGL+VI+AEA IS+WY+YYRENGLV SPGG+PGEDLDVL  LTYSRNL
Sbjct: 354 AYGAATTGVEGLEVILAEAGISSWYNYYRENGLVRSPGGFPGEDLDVLAALTYSRNLDGA 413

Query: 413 DYIKNNDCYQALLNEQSKAIDRQSGDYNQYWHDRNYLTHVNNVKSRVVYTHGLQDWNVKP 472
           D++K N  Y+  L E + A+DR+SGDYNQ+WHDRNYL + + VK+ V+  HGLQDWNV P
Sbjct: 414 DFLKGNAEYEKRLAEMTAALDRKSGDYNQFWHDRNYLINTDKVKADVLIVHGLQDWNVTP 473

Query: 473 RHVYKVFNALPQTIKKHLFLHQGQHVYMHNWQSIDFRESMNALLSQELLGIDNHFQLEEV 532
              Y  + ALP+   KH FLH+G H+YM++WQSIDF E++NA    +LL  D +  L  V
Sbjct: 474 EQAYNFWKALPEGHAKHAFLHRGAHIYMNSWQSIDFSETINAYFVAKLLDRDLNLNLPPV 533

Query: 533 IWQDNTTEQTWQVLDAFGGNHQEQIGLGD---SKKLIDNHYDKEAFDTYCKDFNVFKNDL 589
           I Q+N+ +Q W +++ FG N Q ++ LG    S    DNHYD E F  Y KDFNVFK DL
Sbjct: 534 ILQENSKDQVWTMMNDFGANTQIKLPLGKTAVSFAQFDNHYDDETFKKYSKDFNVFKKDL 593

Query: 590 FKGNNKTNQITINLPLKKNYLLNGQCKLHLRVKTSDKKAILSAQILDYGPKKRFKDTPTI 649
           F+  NK N+  I+L L     +NG  +L LR+K +D K  LSAQILD+G KKR +D   +
Sbjct: 594 FE--NKANEAVIDLELPSMLTINGPVELELRLKLNDTKGFLSAQILDFGQKKRLEDKARV 651

Query: 650 KFLNSLDNGKNFAREALRELPFTKDHYRVISKGVLNLQNRTDLLTIEAIEPEQWFDIEFS 709
           K    LD G+NF  + L ELP  +  Y++I+KG  NLQN+ +LLT+  ++ ++WF I+F
Sbjct: 652 KDFKVLDRGRNFMLDDLVELPLVESPYQLITKGFTNLQNQ-NLLTVSDLKADEWFTIKFE 710

Query: 710 LQPSIYQLSKGDNLRIILYTTDFEHTIRDNASYSITVDLSQSYLTIP 756
           LQP+IY L K D LR+ILY+TDFEHT+RDN   +  +DLSQS L IP
Sbjct: 711 LQPTIYHLEKADKLRVILYSTDFEHTVRDNRKVTYEIDLSQSKLIIP 757
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4991> which encodes the amino acid sequence <SEQ ID 4992>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2553(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 481/758 (63%), Positives = 587/758 (76%), Gaps = 4/758 (0%)

Query: 1   MRYNQFSYIPTKPNEAFEELKGLGFPLNKKNSDKANLEAFLRHSFLNQTDTDYALSLLIV 60
           MRYNQFSYIPT    A EELK LGF L+ + + KA+LE+FLR  F +  D+DY LS LI
Sbjct: 1   MRYNQFSYIPTSLERAAEELKELGFDLDLQKTAKASLESFLRKLFFHYPDSDYPLSHLIA 60
```

```
Query: 61  DAKTDALTFFKSNSDLTLENLQWIYLQLLGFIPFVDFKDPKAFLQDINFPVSYDN--IFQ 118
               DAL+FF+S  +L+ E    + LQ+LGFIP VDF +  AFL  + FP+ +D   I +
Sbjct: 61  KNDMDALSFFQSEQELSKEVFDLLALQVLGFIPGVDFTEADAFLDKLAFPIHFDETEIIK 120

Query: 119 SLHHLLACRGKSGNTLIDQLVADGLLHADNHYHFFNGKSLATFNTNQLIREVVYVETSLD 178
            +HHLLA R KSG TLID LV+ G+L  DN YHFFNGKSLATF+T+QLIREVVYVE  LD
Sbjct: 121 HIHHLLATRCKSGMTLIDDLVSQGMLTMDNDYHFFNGKSLATFDTSQLIREVVYVEAPLD 180

Query: 179 TMSSGEHDLVKVNIIRPTTEHTIPTMMTASPYHQGINDPAADQKTYQMEGALAVKQPKHI 238
           T   G+ DL+KVNIIRP ++  +PT+MT SPYHQGIN+ A D+K Y+ME  L VK+ + I
Sbjct: 181 TDQDGQLDLIKVNIIRPQSQKPLPTLMTPSPYHQGINEVANDKKLYRMEKELVVKKRRQI 240

Query: 239 QVDTKPFKEEVKHPSKLPISPATESFTHIDSYSLNDYFLSRGFANIYVSGVGTAGSTGFM 298
            V+ + F      P KLPI    ESF++I+SYSLNDYFL+RGFANIYVSGVGTAGSTGFM
Sbjct: 241 TVEDRDFIPLETQPCKLPIGQNLESFSYINSYSLNDYFLARGFANIYVSGVGTAGSTGFM 300

Query: 299 TSGDYQQIQSFKAVIDWLNGKVTAFTSHKRDKQVKANWSNGLVATTGKSYLGTMSTGLAT 358
           TSG+Y QI+SFKAVIDWLNG+ TA+TSH +  QV+A+W+NGLV TTGKSYLGTMSTGLAT
Sbjct: 301 TSGNYAQIESFKAVIDWLNGRATAYTSHSKTHQVRADWANGLVCTTGKSYLGTMSTGLAT 360

Query: 359 TGVEGLKVIIAEAAISTWYDYYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAGDYIKNN 418
           TGV+GL +IIAE+AIS+WY+YYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAGDY+++N
Sbjct: 361 TGVDGLAMIIAESAISSWYNYYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAGDYLRHN 420

Query: 419 DCYQALLNEQSKAIDRQSGDYNQYWHDRNYLTHVNNVKSRVVYTHGLQDWNVKPRHVYKV 478
           D YQ LLN+QS+A+DRQSGDYNQ+WHDRNYL + + +K  VVYTHGLQDWNVKPR VY++
Sbjct: 421 DRYQELLNQQSQALDRQSGDYNQFWHDRNYLKNAHQIKCDVVYTHGLQDWNVKPRQVYEI 480

Query: 479 FNALPQTIKKHLFLHQGQHVYMHNWQSIDFRESMNALLSQELLGIDNHFQLEEVIWQDNT 538
           FNALP TI KHLFLHQG+HVYMHNWQSIDFRESMNALL Q+LLG+ N F L E+IWQDNT
Sbjct: 481 FNALPSTINKHLFLHQGEHVYMHNWQSIDFRESMNALLCQKLLGLANDFSLPEMIWQDNT 540

Query: 539 TEQTWQVLDAFGGNHQEQIGLGDSKKLIDNHYDKEAFDTYCKDFNVFKNDLFKGNNKTNQ 598
             Q WQ    FG +  +++ LG    LIDNHY ++ F  Y KDF  FK  LFKG  K NQ
Sbjct: 541 CPQNWQERKVFGTSTIKELDLGQELLLIDNHYGEDEFKAYGKDFRAFKAALFKG--KANQ 598

Query: 599 ITINLPLKKNYLLNGQCKLHLRVKTSDKKAILSAQILDYGPKKRFKDTPTIKFLNSLDNG 658
             I++ L+++  +NG+  L L+VK+S+ K +LSAQILDYG KKR  D P     +S+DNG
Sbjct: 599 ALIDILLEEDLPINGEIVLQLKVKSSENKGLLSAQILDYGKKKRLGDLPIALTQSSIDNG 658

Query: 659 KNFAREALRELPFTKDHYRVISKGVLNLQNRTDLLTIEAIEPEQWFDIEFSLQPSIYQLS 718
           +NF+RE L+ELPF +D YRVISKG +NLQNR +L +IE I   +W  +   LQP+IY L
Sbjct: 659 QNFSREPLKELPFREDSYRVISKGFMNLQNRNNLSSIETIPNNKWMTVRLPLQPTIYHLE 718

Query: 719 KGDNLRIILYTTDFEHTIRDNASYSITVDLSQSYLTIP 756
           KGD LR+ILYTTDFEHT+RDN++Y++T+DLSQS L +P
Sbjct: 719 KGDTLRVILYTTDFEHTVRDNSNYALTIDLSQSQLIVP 756
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1618

A DNA sequence (GBSx1713) was identified in *S.agalactiae* <SEQ ID 4993> which encodes the amino acid sequence <SEQ ID 4994>. This protein is predicted to be PrfA. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3976(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10125> which encodes amino acid sequence <SEQ ID 10126> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA65740 GB:X97014 PrfA [Listeria seeligeri]
 Identities = 54/181 (29%), Positives = 95/181 (51%), Gaps = 1/181 (0%)

Query: 38  DYTYILKDGIVKQSVLSKYGTEFNLRYVTGLEITSILNTDYSQHMGEPYNVRIESETAHF 97
            +Y   L +G+ K + +S+ G   NL+Y  G  I      D  + +G  YN+ + SE A
Sbjct: 36  EYCIFLHEGVAKLTSISESGDILNLQYYKGAFIIMTGFIDTEKSLGY-YNLEVVSEQAAA 94

Query: 98  YKVRRSTFLKDINNDIELQGYVKDFYHNRLEKSMKKMQCMLTNGRIGAISTQLYDLSKMF 157
            Y ++ S   + ++ D++   Y+ D    ++  S+ K     +NG++G+I  Q   L+ ++
Sbjct: 95  YIIKISDLKELVSKDLKQLFYIIDTLQKQVSYSLAKFNDFSSNGKVGSICGQFLILAYVY 154

Query: 158 GEERDNGDIYINFVITNEELGKFCGISTGSSVSRILKQLKDDHIIRIEKQHIIITNVEKLK 218
            GEE  NG       +T +ELG   GI+  S+VSRI+ +LK +++I  +  +  I N+  LK
Sbjct: 155 GEETPNGIKITLEKLTMQELGCSSGIAHSSAVSRIISKLKQENVIEYKDSYFYIKNIAYLK 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4995> which encodes the amino acid sequence <SEQ ID 4996>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4088(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 186/223 (83%), Positives = 203/223 (90%)

Query: 1   MEEVMNHQILQNYINSHNLPIIEKDYHKYLTFESLEEDYTYILKDGIVKQSVLSKYGTEF 60
            +E+ +NH ILQ YI++HN PIIEK YHKYLTFESLEED+TYILKDGIVKQSVLSKYG EF
Sbjct: 17  LEKSVNHHILQRYIDNHNFPIIEKSYHKYLTFESLEEDFTYILKDGIVKQSVLSKYGMEF 76

Query: 61  NLRYVTGLEITSILNTDYSQHMGEPYNVRIESETAHFYKVRRSTFLKDINNDIELQGYVK 120
            NLRYVTGLEITS+LNT YS+ MGEPYNVRIESE A FYKVRRS FLKDIN DIELQGYVK
Sbjct: 77  NLRYVTGLEITSVLNTGYSKDMGEPYNVRIESEKASFYKVRRSAFLKDINEDIELQGYVK 136

Query: 121 DFYHNRLEKSMKKMQCMLTNGRIGAISTQLYDLSKMFGEERDNGDIYINFVITNEELGKF 180
            DFYHNRL+KSMKKMQCMLTNGRIGAISTQ+YDL  +FGEE  NG I INFVITNEELGKF
Sbjct: 137 DFYHNRLQKSMKKMQCMLTNGRIGAISTQIYDLMTLFGEELPNGQILINFVITNEELGKF 196

Query: 181 CGISTGSSVSRILKQLKDDHIIRIEKQHIIITNVEKLKDHIVF 223
            CGIST SSVSRILKQLK+ +IIRI+KQHIIITN++KLKD+IVF
Sbjct: 197 CGISTASSVSRILKQLKEKNIIRIDKQHIIITNLDKLKDNIVF 239
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1619

A DNA sequence (GBSx1714) was identified in *S.agalactiae* <SEQ ID 4997> which encodes the amino acid sequence <SEQ ID 4998>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.33   Transmembrane  167 - 183 ( 159 - 193)
   INTEGRAL    Likelihood = -7.96   Transmembrane   18 -  34 (  10 -  37)
   INTEGRAL    Likelihood = -7.75   Transmembrane  373 - 389 ( 369 - 392)
```

```
    INTEGRAL    Likelihood = -5.68   Transmembrane  214 - 230 ( 212 - 234)
    INTEGRAL    Likelihood = -4.78   Transmembrane  243 - 259 ( 241 - 262)
    INTEGRAL    Likelihood = -2.71   Transmembrane   48 -  64 (  47 -  65)
    INTEGRAL    Likelihood = -2.60   Transmembrane  283 - 299 ( 283 - 300)

----- Final Results -----
               bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15662 GB:Z99122 similar to antibiotic resistance protein
          [Bacillus subtilis]
 Identities = 106/401 (26%), Positives = 199/401 (49%), Gaps = 21/401 (5%)

Query: 3   DKLFNKHFIGITILNFIVYMVYYLFTVIIAFIATKELGVSTSQAGLATGIYIVGTLIARL 62
            D ++ K FI + ++N  V++ +Y F  ++     +ELG + SQ GL   ++++  +I R
Sbjct: 5   DAIWTKDFIMVLLVNLFVFVFFYTFLTVLPIYTLQELGGTESQGGLLISLFLLSAIITRP 64

Query: 63  IFGKQLEVLGRKLVLRGGAIFYLLTTLAYFYMPSIGVMYLVRFLNGFGYGVVSTATNTIV 122
              G  +E  G+K +       + L++  Y  + +  ++  +RF  G  + +++T T  I
Sbjct: 65  FSGAIVERFGKKRMAIVSMALFALSSFLYMPIHNFSLLLGLRFFQGIWFSILTTVTGAIA 124

Query: 123 TAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHINFKMVIVLCSILIAIVVLGA 182
               IPA +RGEG+ ++ +S +LA AIGPF+G  ++    ++F +     ++ +   +L +
Sbjct: 125 ADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLMRV--VSFPVFFTAFALFMVAGLLVS 182

Query: 183 FVFPVKNITLNPEQLAKSKSWTIDSF-----IEKKAIFITIIAFLMGISYASVLGFQKLY 237
            F+  V          +K    T+  F      EK A+ I  +   +   Y++V  +  ++
Sbjct: 183 FLIKVPQ--------SKDSGTTVFRFAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVF 234

Query: 238 TTEINLMTVGAYFFIVYALVITLTRPSMGRLMDAKGDKWVLYPSYLFLTLGLALLGSAMG 297
               ++L  +  YFF+ +A+ + + RP  G+L D  G   V+YPS L  ++GL +L
Sbjct: 235 AKSVDLSDISGYFFVCFAVTMMIARPFTGKLFDKVGPGIVIYPSILIFSVGLCMLSFTHS 294

Query: 298 SVTYLLSGALIGFGYGTFMSCGQAASIKGVEEHRFNTAMSTYMIGLDLGLGAGPYILGLV 357
             +  LLSGA+IG GYG+ + C Q  +I+    HR   A +T+    D G+  G Y+ GL
Sbjct: 295 GLMLLLSGAVIGLGYGSIVPCMQTLAIQKSPAHRSGFATATFFTFFDSGIAVGSYVFGL- 353

Query: 358 KDGFLGAGVQSFRELFWIAAIIPVVCGILYFLKSSRQVETK 398
               F+ +    F  ++  A +  ++  +LY     +  E +
Sbjct: 354 ---FVASA--GFSAIYLTAGLFVLIALLLYTWSQKKPAEAE 389
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4999> which encodes the amino acid sequence <SEQ ID 5000>. Analysis of this protein sequence reveals the following:

```
      Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-12.31   Transmembrane  202 - 218 ( 194 - 225)
    INTEGRAL    Likelihood = -7.80   Transmembrane   53 -  69 (  44 -  71)
    INTEGRAL    Likelihood = -7.17   Transmembrane  407 - 423 ( 404 - 426)
    INTEGRAL    Likelihood = -5.26   Transmembrane  249 - 265 ( 247 - 269)
    INTEGRAL    Likelihood = -3.77   Transmembrane  279 - 295 ( 276 - 297)
    INTEGRAL    Likelihood = -2.23   Transmembrane   11 -  27 (  10 -  27)
    INTEGRAL    Likelihood = -2.13   Transmembrane   83 -  99 (  82 -  99)
    INTEGRAL    Likelihood = -1.91   Transmembrane  312 - 328 ( 311 - 328)

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15662 GB:Z99122 similar to antibiotic resistance protein
          [Bacillus subtilis]
```

```
 Identities = 110/390 (28%), Positives = 194/390 (49%), Gaps = 11/390 (2%)

Query: 38  EKLFNKHFVAITVINFIVYMVYYLFTVIIAFVATRELGAQTSQAGLATGIYILGTLLARL 97
            + ++ K F+ + ++N  V++ +Y F  ++     +ELG   SQ GL   +++L  ++ R
Sbjct: 5   DAIWTKDFIMVLLVNLFVFVFFYTFLTVLPIYTLQELGGTESQGGLLISLFLLSAIITRP 64

Query: 98  IFGKQLEVFGRRLVLRGGAIFYLLTTLAYFYMPTISMMYLVRFLNGFGYGVVSTATNTIV 157
              G  +E FG++ +       + L++  Y  +   S++  +RF  G  + +++T T  I
Sbjct: 65  FSGAIVERFGKKRMAIVSMALFALSSFLYMPIHNFSLLLGLRFFQGIWFSILTTVTGAIA 124

Query: 158 TAYIPARKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHIDFRMIIVLCSVLIGCVVVGA 217
               IPA++RGEG+ ++ +S +LA AIGPF+G  ++    + F +     ++ +   ++ +
Sbjct: 125 ADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLMRV--VSFPVFFTAFALFMVAGLLVS 182

Query: 218 FAFPVKNMSLNAEQLAKTKSWTVDSFIEKKALFITAIAFLMGIAYASVLGFQKLYTSEIH 277
            F   V     +   + +   +      EK AL I  +   +   Y++V  + ++    +
Sbjct: 183 FLIKVPQSKDSGTTVFR---FAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVFAKSVD 239

Query: 278 LTTVGAYFFVVYALIITITRPAMGRLMDAKGDKWVLYPSYLFLAMGLFLLGSVSSGGSYL 337
            L+ +  YFFV +A+ + I RP  G+L D  G   V+YPS L  ++GL +L    SG   L
Sbjct: 240 LSDISGYFFVCFAVTMMIARPFTGKLFDKVGPGIVIYPSILIFSVGLCMLSFTHSGLMLL 299

Query: 338 LSGALIGFGYGTFMSCGQAASIQGVDEHRFNTAMSTYMIGLDLGLGAGPYLLGLIKDLAL 397
            LSGA+IG GYG+ + C Q  +IQ    HR   A +T+    D G+  G Y+ GL
Sbjct: 300 LSGAVIGLGYGSIVPCMQTLAIQKSPAHRSGFATATFFTFFDSGIAVGSYVFGLF----- 354

Query: 398 GSGVASFRHLFWLAAVIPLICTLLYLLKTK 427
                A F  ++  A +  LI  LLY    K
Sbjct: 355 -VASAGFSAIYLTAGLFVLIALLLYTWSQK 383
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 328/396 (82%), Positives = 370/396 (92%), Gaps = 1/396 (0%)

Query: 1   MEDKLFNKHFIGITILNFIVYMVYYLFTVIIAFIATKELGVSTSQAGLATGIYIVGTLIA 60
            ME+KLFNKHF+ IT++NFIVYMVYYLFTVIIAF+AT+ELG  TSQAGLATGIYI+GTL+A
Sbjct: 36  MEEKLFNKHFVAITVINFIVYMVYYLFTVIIAFVATRELGAQTSQAGLATGIYILGTLLA 95

Query: 61  RLIFGKQLEVLGRKLVLRGGAIFYLLTTLAYFYMPSIGVMYLVRFLNGFGYGVVSTATNT 120
            RLIFGKQLEV GR+LVLRGGAIFYLLTTLAYFYMP+I +MYLVRFLNGFGYGVVSTATNT
Sbjct: 96  RLIFGKQLEVFGRRLVLRGGAIFYLLTTLAYFYMPTISMMYLVRFLNGFGYGVVSTATNT 155

Query: 121 IVTAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHINFKMVIVLCSILIAIVVL 180
            IVTAYIPA KRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHI+F+M+IVLCS+LI  VV+
Sbjct: 156 IVTAYIPARKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHIDFRMIIVLCSVLIGCVVV 215

Query: 181 GAFVFPVKNITLNPEQLAKSKSWTIDSFIEKKAIFITIIAFLMGISYASVLGFQKLYTTE 240
            GAF FPVKN++LN EQLAK+KSWT+DSFIEKKA+FIT IAFLMGI+YASVLGFQKLYT+E
Sbjct: 216 GAFAFPVKNMSLNAEQLAKTKSWTVDSFIEKKALFITAIAFLMGIAYASVLGFQKLYTSE 275

Query: 241 INLMTVGAYFFIVYALVITLTRPSMGRLMDAKGDKWVLYPSYLFLTLGLALLGSAMGSVT 300
            I+L TVGAYFF+VYAL+IT+TRP+MGRLMDAKGDKWVLYPSYLFL +GL LLGS     +
Sbjct: 276 IHLTTVGAYFFVVYALIITITRPAMGRLMDAKGDKWVLYPSYLFLAMGLFLLGSVSSGGS 335

Query: 301 YLLSGALIGFGYGTFMSCGQAASIKGVEEHRFNTAMSTYMIGLDLGLGAGPYILGLVKDG 360
            YLLSGALIGFGYGTFMSCGQAASI+GV+EHRFNTAMSTYMIGLDLGLGAGPY+LGL+KD
Sbjct: 336 YLLSGALIGFGYGTFMSCGQAASIQGVDEHRFNTAMSTYMIGLDLGLGAGPYLLGLIKDL 395

Query: 361 FLGAGVQSFRELFWIAAIIPVVCGILYFLKS-SRQV 395
             LG+GV SFR LFW+AA+IP++C +LY LK+ +RQV
Sbjct: 396 ALGSGVASFRHLFWLAAVIPLICTLLYLLKTKTRQV 431
```

A related GBS gene <SEQ ID 8863> and protein <SEQ ID 8864> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      8.26
```

```
GvH: Signal Score (-7.5): -5.21
     Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 7 value: -14.33 threshold:  0.0
   INTEGRAL    Likelihood =-14.33   Transmembrane  167 - 183 ( 159 - 193)
   INTEGRAL    Likelihood = -7.96   Transmembrane   18 -  34 (  10 -  37)
   INTEGRAL    Likelihood = -7.75   Transmembrane  373 - 389 ( 369 - 392)
   INTEGRAL    Likelihood = -5.68   Transmembrane  214 - 230 ( 212 - 234)
   INTEGRAL    Likelihood = -4.78   Transmembrane  243 - 259 ( 241 - 262)
   INTEGRAL    Likelihood = -2.71   Transmembrane   48 -  64 (  47 -  65)
   INTEGRAL    Likelihood = -2.60   Transmembrane  283 - 299 ( 283 - 300)
   PERIPHERAL  Likelihood =  0.69     341
 modified ALOM score:   3.37

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01003(307 - 1494 of 1800)
EGAD|108032|BS3640(5 - 389 of 396) hypothetical protein {Bacillus subtilis}
GP|1684651|emb|CAB05383.1||Z82987 unknown similar to quinolon resistance protein NorA
{Bacillus subtilis} GP|2636170|emb|CAB15662.1||Z99122 similar to antibiotic resistance
protein {Bacillus subtilis} PIR|B70065|B70065 antibiotic resistance protein homolog ywoG -
Bacillus subtilis
%Match = 14.9
%Identity = 26.3  %Similarity = 53.4
Matches = 102  Mismatches = 178  Conservative Sub.s = 105

204       234       264       294       324       354       384       414
TTLTFVNAV*Y*HLYYTIEISYLLIFL*NVYENEIEKKEPFALEDKLFNKHFIGITILNFIVYMVYYLFTVIIAFIATKE
                                            | :: | || : ::|: |:: :| |  ::     :|
                                        MKKADAIWTKDFIMVLLVNLFVFVFFYTFLTVLPIYTLQE
                                                10        20        30        40

444       474       504       534       564       594       624       654
LGVSTSQAGLATGIYIVGTLIARLIFGKQLEVLGRKLVLRGGAIFYLLTTLAYFYMPSIGVMYLVRFLNGFGYGVVSTAT
|| : || ||   ::::  :| |   |  :| :|:| :      :: |::: |  : :  ::  :||: |  : :::| |
LGGTESQGGLLISLFLLSAIITRPFSGAIVERFGKKRMAIVSMALFALSSFLYMPIHNFSLLLGLRFFQGIWFSILTTVT
        50        60        70        80        90       100       110       120

684       714       744       774       804       834       864       894
NTIVTAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHINFKMVIVLCSILIAIVVLGAFVFPVKNITLNPEQLA
  |    ||| :||||: :: :| :|| |||||:|  ::    ::| :      ::::  :| :|:  |     :   :
GAIAADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLM--RVVSFPVFFTAFALFMVAGLLVSFLIKVPQSKDSGTTVF
       130       140       150         160       170       180       190

924       954       984       1014      1044      1074      1104      1134
KSKSWTIDSFIEKKAIFITIIAFLMGISYASVLGFQKLYTTEINLMTVGAYFFIVYALVITLTRPSMGRLMDAKGDKWVL
:   :      || |: |  : :::   |::|  :  ::   ::|  :  |||: :|: : : ||  |:| |  |   |:
R---FAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVFAKSVDLSDISGYFFVCFAVTMMIARPFTGKLFDKVGPGIVI
            210       220       230       240       250       260       270

1164      1194      1224      1254      1284      1314      1344      1374
YPSYLFLTLGLALLGSAMGSVTYLLSGALIGFGYGTFMSCGQAASIKGVEEHRFNTAMSTYMIGLDLGLGAGPYILGLVK
||| | :::|| :|      :  |||||:||:|||: : | |  :|:    ||   | :|:   :| |:  | |::||
YPSILIFSVGLCMLSFTHSGLMLLLSGAVIGLGYGSIVPCMQTLAIQKSPAHRSGFATATFFTFFDSGIAVGSYVFGL--
            290       300       310       320       330       340       350

1404      1434      1464      1494      1524      1554      1584      1614
DGFLGAGVQSFRELFWIAAIIPVVCGILYFLKSSRQVETKTI*KGGIKL*HKNMSVFLLLLMGLTSQNWR*KKG*MLLFV
          |  ::  | :  ::  :||     :  | :
----FVASAGFSAIYLTAGLFVLIALLLYTWSQKKPAEAEGKVSIAE
        360       370       380       390
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1620

A DNA sequence (GBSx1715) was identified in *S.agalactiae* <SEQ ID 5001> which encodes the amino acid sequence <SEQ ID 5002>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06903 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 52/143 (36%), Positives = 84/143 (58%)

Query: 5   YERILIAIDGSYESELAVEKGINVALRNDAELLLTHVIDAHAYQSEGVFSDYVFDRQEQE 64
           Y  IL+A+DGS +++ A+ K  N A    A+L + HVID+ ++ +   +   V    E +
Sbjct: 2   YNHILVAVDGSTQAKRALYKAFNYAKEFKADLFICHVIDSRSFATVEQYDRTVVGAAELD 61

Query: 65  SADVLAYFEKLAHSKGLTKIKKITEIGNPKTLLAKDIPIREKADLIMVGATGLNTFERLL 124
              +L  + + A   G+ K+  I + G+PK  ++K I  +   DLI+ GATGLN  ER L
Sbjct: 62  GKKLLQRYSEEAEKAGVDKVHTILDFGSPKANISKTIAQKYDIDLIITGATGLNAVERFL 121

Query: 125 IGSTSEYILRHSKVDMLVVRDSK 147
           +GS SE + RH+K D+L+VR+ +
Sbjct: 122 MGSVSESVARHAKCDVLIVRNDQ 144
```

There is also homology to SEQ ID 3658:

```
Identities = 105/150 (70%), Positives = 121/150 (80%)

Query: 1   MTQKYERILIAIDGSYESELAVEKGINVALRNDAELLLTHVIDAHAYQSEGVFSDYVFDR 60
           M+ KY+RIL+AIDGSYESELA  KG+NVALRNDA LLL HVID  A QS   F  Y++++
Sbjct: 31  MSLKYKRILVAIDGSYESELAFNKGVNVALRNDATLLLVHVIDTRALQSVATFDTYIYEK 90

Query: 61  QEQESADVLAYFEKLAHSKGLTKIKKITEIGNPKTLLAKDIPIREKADLIMVGATGLNTF 120
            EQE+ DVL  FEK A   G+T IK+I E GNPK LLA DIP RE ADLIMVGATGLNTF
Sbjct: 91  LEQEAKDVLDDFEKQAQIAGITNIKQIIEFGNPKNLLAHDIPDRENADLIMVGATGLNTF 150

Query: 121 ERLLIGSTSEYILRHSKVDMLVVRDSKKTL 150
           ERLLIGS+SEYI+RH+K+D+LVVRDS KTL
Sbjct: 151 ERLLIGSSSEYIMRHAKIDLLVVRDSTKTL 180
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1621

A DNA sequence (GBSx1716) was identified in *S.agalactiae* <SEQ ID 5003> which encodes the amino acid sequence <SEQ ID 5004>. This protein is predicted to be glycerol uptake facilitator protein (glpF). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.65   Transmembrane  261 - 277 ( 257 - 281)
   INTEGRAL    Likelihood = -5.73   Transmembrane  201 - 217 ( 199 - 222)
```

```
   INTEGRAL    Likelihood = -4.51   Transmembrane   92 - 108 (  91 - 110)
   INTEGRAL    Likelihood = -4.30   Transmembrane   44 -  60 (  42 -  62)
   INTEGRAL    Likelihood = -2.18   Transmembrane   15 -  31 (  11 -  31)
   INTEGRAL    Likelihood = -1.54   Transmembrane  150 - 166 ( 149 - 166)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25231 GB:M58315 putative [Lactococcus lactis]
 Identities = 183/290 (63%), Positives = 228/290 (78%), Gaps = 10/290 (3%)

Query: 2   IEITWTVKYITEFIATAFLIIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVMMPALMF 61
            +++TWTVKYITEF+ TA LII+GNGAVANV+LKGTK +   W+II  GYGLGVM+PA+ F
Sbjct: 1   MDVTWTVKYITEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVMLPAVAF 60

Query: 62  GNVSGNHINPAFTLGLAFSGLFPWAHVGQYILAQILGAMFGQLVVVMVYQPYFVKTENPN 121
            GN++ + INPAFTLGLA SGLFPWAHV QYI+AQ+LGAMFGQL++VMVY+PY++KT+NPN
Sbjct: 61  GNIT-SQINPAFTLGLAASGLFPWAHVAQYIIAQVLGAMFGQLLIVMVYRPYYLKTQNPN 119

Query: 122 HVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSFVLFFGALALTKNYFGVE----LVG 177
             +LG+FSTI  +DD  + +R  A INGFLNEF+GSFVLFFGA+A T  +FG +    +
Sbjct: 120 AILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIFFGSQSITWMTN 179

Query: 178 KLVQAGYDQTTAATRISPYVTGSLA-----VAHLGIGFLVMTLVASLGGPTGPALNPARD 232
             L   G D +++       +V  S A     +AHL +GFLVM LV +LGGPTGP LNPARD
Sbjct: 180 YLKGQGADVSSSDVMNQIWVQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGLNPARD 239

Query: 233 LGPRIVHRLLPKQILGQAKEDSKWWYAWVPVLAPIVASILAVALFKLLYL 282
             GPR+VH LLPK +LG+AK  SKWWYAWVPVLAPI+AS+ AVALFK++YL
Sbjct: 240 FGPRLVHSLLPKSVLGEAKGSSKWWYAWVPVLAPILASLAAVALFKMIYL 289
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5005> which encodes the amino acid sequence <SEQ ID 5006>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.18   Transmembrane  293 - 309 ( 288 - 314)
   INTEGRAL    Likelihood = -7.43   Transmembrane    2 -  18 (   1 -  20)
   INTEGRAL    Likelihood = -7.38   Transmembrane  233 - 249 ( 228 - 256)
   INTEGRAL    Likelihood = -5.57   Transmembrane  124 - 140 ( 123 - 142)
   INTEGRAL    Likelihood = -2.87   Transmembrane   76 -  92 (  75 -  93)
   INTEGRAL    Likelihood = -2.18   Transmembrane   47 -  63 (  43 -  63)
   INTEGRAL    Likelihood = -1.54   Transmembrane  182 - 198 ( 181 - 198)

----- Final Results -----
               bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA25231 GB:M58315 putative [Lactococcus lactis]
 Identities = 176/290 (60%), Positives = 228/290 (77%), Gaps = 10/290 (3%)

Query: 34  MEMTWTVKYITEFIATAFLIIILGNGAVANVDLKGTKGHNSGWLVIAFGYGLGVMMPALMF 93
            M++TWTVKYITEF+ TA LII+GNGAVANV+LKGTK H   W++I +GYGLGVM+PA+ F
Sbjct: 1   MDVTWTVKYITEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVMLPAVAF 60

Query: 94  GNVSGNHINPAFTVGLAVSGLFPWAHVLQYVVAQLLGAIFGQLVVVMVYKPYFMKTENPN 153
            GN++ + INPAFT+GLA SGLFPWAHV QY++AQ+LGA+FGQL++VMVY+PY++KT+NPN
Sbjct: 61  GNIT-SQINPAFTLGLAASGLFPWAHVAQYIIAQVLGAMFGQLLIVMVYRPYYLKTQNPN 119

Query: 154 HVLGSFSTISSLDNGQKDSHKASYINGFLNEFVGSFVLFFGALALTKNYFGVELVGKLIE 213
```

```
            +LG+FSTI ++D+  + +   + INGFLNEF+GSFVLFFGA+A T  +FG + +  +
Sbjct: 120 AILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIFFGSQSITWMTN 179

Query: 214 ------AGYDQTTAATQISPYVTGSLA---VAHIGIGFLVMVLVTSLGGPTGPALNPARD 264
                  A    +    QI    +G+ A   +AH+ +GFLVM LV +LGGPTGP LNPARD
Sbjct: 180 YLKGQGADVSSSDVMNQIWVQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGLNPARD 239

Query: 265 FGPRLLHHFLPKSVLGQAKGDSKWWYAWVPVVAPILAAIVAVAAFKYLYI 314
            FGPRL+H  LPKSVLG+AKG SKWWYAWVPV+APILA++ AVA FK +Y+
Sbjct: 240 FGPRLVHSLLPKSVLGEAKGSSKWWYAWVPVLAPILASLAAVALFKMIYL 289
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 240/281 (85%), Positives = 267/281 (94%)

Query: 2   IEITWTVKYITEFIATAFLIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVMMPALMF 61
            +E+TWTVKYITEFIATAFLIILGNGAVANVDLKGTKG+NSGW++IA GYGLGVMMPALMF
Sbjct: 34  MEMTWTVKYITEFIATAFLIILGNGAVANVDLKGTKGHNSGWLVIAFGYGLGVMMPALMF 93

Query: 62  GNVSGNHINPAFTLGLAFSGLFPWAHVGQYILAQILGAMFGQLVVVMVYQPYFVKTENPN 121
            GNVSGNHINPAFT+GLA SGLFPWAHV QY++AQ+LGA+FGQLVVVMVY+PYF+KTENPN
Sbjct: 94  GNVSGNHINPAFTVGLAVSGLFPWAHVLQYVVAQLLGAIFGQLVVVMVYKPYFMKTENPN 153

Query: 122 HVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSFVLFFGALALTKNYFGVELVGKLVQ 181
            HVLGSFSTIS+LD+GQK S KA+YINGFLNEFVGSFVLFFGALALTKNYFGVELVGKL++
Sbjct: 154 HVLGSFSTISSLDNGQKDSHKASYINGFLNEFVGSFVLFFGALALTKNYFGVELVGKLIE 213

Query: 182 AGYDQTTAATRISPYVTGSLAVAHLGIGFLVMTLVASLGGPTGPALNPARDLGPRIVHRL 241
            AGYDQTTAAT+ISPYVTGSLAVAH+GIGFLVM LV SLGGPTGPALNPARD GPR++H
Sbjct: 214 AGYDQTTAATQISPYVTGSLAVAHIGIGFLVMVLVTSLGGPTGPALNPARDFGPRLLHHF 273

Query: 242 LPKQILGQAKEDSKWWYAWVPVLAPIVASILAVALFKLLYL 282
            LPK +LGQAK DSKWWYAWVPV+API+A+I+AVA FK LY+
Sbjct: 274 LPKSVLGQAKGDSKWWYAWVPVVAPILAAIVAVAAFKYLYI 314
```

A related GBS gene <SEQ ID 8865> and protein <SEQ ID 8866> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      2.81
GvH: Signal Score (-7.5): -3.6
     Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 6 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  261 - 277 ( 257 - 281)
   INTEGRAL    Likelihood = -5.73   Transmembrane  201 - 217 ( 199 - 222)
   INTEGRAL    Likelihood = -4.51   Transmembrane   92 - 108 (  91 - 110)
   INTEGRAL    Likelihood = -4.30   Transmembrane   44 -  60 (  42 -  62)
   INTEGRAL    Likelihood = -2.18   Transmembrane   15 -  31 (  11 -  31)
   INTEGRAL    Likelihood = -1.54   Transmembrane  150 - 166 ( 149 - 166)
   PERIPHERAL  Likelihood =  2.92      72
 modified ALOM score:   2.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01006(304 - 1146 of 1446)
EGAD|14239|14211(1 - 289 of 289) hypothetical 30.9 kd protein in pepx 5'region {Lactococcus
lactis} SP|P22094|YDP1_LACLC HYPOTHETICAL 30.9 KDA PROTEIN IN PEPX 5'REGION (ORF1).
GP|455286|gb|AAA25206.1||M35865    ORF1    (put.);    putative    {Lactococcus    lactis}
GP|149527|gb|AAA25231.1||M58315    putative    {Lactococcus    lactis}    PIR|B43747|B43747
```

```
hypothetical protein (pepXP 5' region) - Lactococcus lactis subsp. cremoris
PIR|B43748|B43748 hypothetical protein (pepX 5' region) - Lactococcus lactis subsp. lactis
%Match = 37.5
%Identity = 64.4  %Similarity = 81.3
Matches = 183  Mismatches = 49  Conservative Sub.s = 48

123       153       183       213       243       273       303       333
*YASRS***ENLIN*IK*STR*SEPSTLFFIKYIWLKILLILFCDKLYNIKLTW*NG*CCKYFFGRKQGLIEITWTVKYI
                                                                      :::|||||||
                                                                      MDVTWTVKYI
                                                                              10

363       393       423       453       483       513       543       573
TEFIATAFLIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVMMPALMFGNVSGNHINPAFTLGLAFSGLFPWAHVGQY
|||: ||:|||:||||||||:||||| :   |:||  |||||||:||: |||:: ::|||||||||| ||||||||| ||
TEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVMLPAVAFGNIT-SQINPAFTLGLAASGLFPWAHVAQY
        20        30        40        50        60         70        80

603       633       663       693       723       753       783       813
ILAQILGAMFGQLVVVMVYQPYFVKTENPNHVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSFVLFFGALALTKNYF
|:||:||||||||::||||:||::||:||| :||:||||  :||  : :|  | ||||||||:|||||||||:| |  :|
IIAQVLGAMFGQLLIVMVYRPYYLKTQNPNAILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIFF
        100       110       120       130       140       150       160

831       861       885       906       936       966       996       1026
G----VELVGKLVQAGYDQTTA--ATRISPYVTG---SLAVAHLGIGFLVMTLVASLGGPTGPALNPARDLGPRIVHRLL
|      :   |   | | :::    :|    :|   |  :||| :||||| || :||||||| ||||||:|||:|| ||
GSQSITWMTNYLKGQGADVSSSDVMNQIWVQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGLNPARDFGPRLVHSLL
        180       190       200       210       220       230       240

1056      1086      1116      1146      1176      1206      1236      1266
PKQILGQAKEDSKWWYAWVPVLAPIVASILAVALFKLLYL**LKKDRFTGLFLF*I*KSASLAS*FRLMMTFGHSFFKGR
|| :||:||  ||||||||||||||:||: ||||||::||
PKSVLGEAKGSSKWWYAWVPVLAPILASLAAVALFKMIYL
        260       270       280
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1622

A DNA sequence (GBSx1717) was identified in *S.agalactiae* <SEQ ID 5007> which encodes the amino acid sequence <SEQ ID 5008>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70    Transmembrane  266 - 282 ( 262 - 290)
   INTEGRAL    Likelihood = -7.96    Transmembrane   25 -  41 (  24 -  50)
   INTEGRAL    Likelihood = -6.42    Transmembrane  110 - 126 ( 105 - 140)
   INTEGRAL    Likelihood = -6.26    Transmembrane  194 - 210 ( 190 - 215)
   INTEGRAL    Likelihood = -5.47    Transmembrane  290 - 306 ( 289 - 310)
   INTEGRAL    Likelihood = -4.35    Transmembrane  128 - 144 ( 127 - 147)
   INTEGRAL    Likelihood = -3.29    Transmembrane  157 - 173 ( 156 - 174)
   INTEGRAL    Likelihood = -2.76    Transmembrane  221 - 237 ( 221 - 240)

----- Final Results -----
               bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9177> which encodes the amino acid sequence <SEQ ID 9178>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 21
>>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood =-10.77   Transmembrane  139 - 155 ( 133 - 161)
     INTEGRAL    Likelihood = -8.28   Transmembrane  245 - 261 ( 240 - 269)
     INTEGRAL    Likelihood = -7.48   Transmembrane  269 - 285 ( 263 - 289)
     INTEGRAL    Likelihood = -7.06   Transmembrane   97 - 113 (  83 - 125)
     INTEGRAL    Likelihood = -6.10   Transmembrane  173 - 189 ( 169 - 194)
     INTEGRAL    Likelihood = -1.44   Transmembrane  200 - 216 ( 200 - 217)

----- Final Results -----
              bacterial membrane --- Certainty=0.531(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 225/301 (74%), Positives = 263/301 (86%)

Query: 10  LTVSLFFCRLDIMNETLLLHGIQLILIIAMIITFYQIVRHIRSQKINPFKRFFTGLWIGF 69
            LT  +FFC+L  MNE L+L  IQ +L+ AM+  F+ +V+H++  KINPFKRF+TG WIG
Sbjct: 1   LTAKVFFCKLVFMNEMLILRLIQALLVSAMLFIFFMLVKHLKKNKINPFKRFWTGFWIGL 60

Query: 70  VTDALDTLGIGSFATTTTFFKLTKLVEDDRKIPATMTAAHVLPVLLQSLCFIFVVKVEAL 129
            +TDALDTLGIGSFATTTT FKLTKLV DDR++P TMT AHVLPVL+QSLCFIFVVKVE L
Sbjct: 61  LTDALDTLGIGSFATTTTCFKLTKLVTDDRQLPGTMTVAHVLPVLIQSLCFIFVVKVEVL 120

Query: 130 TLITMAGAAFIGAFVGAKMTKNWHAPTVQRILGTLLITAAIIMLYRMITNPGAGISDSVH 189
            TL+ MA AAFIGA+ G  +TKNWHAPTVQRILG+LLI AAIIM+ R+I +PG  +SD++H
Sbjct: 121 TLLAMAAAAFIGAYFGTHITKNWHAPTVQRILGSLLIIAAIIMIIRIIYHPGEHLSDTIH 180

Query: 190 GLHGIWLFVGIGFNFIIGVLMTMGLGNYAPELIFFSLMGLSPAVAMPVMMLDAAMIMTAS 249
            GLHGIWLFVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSP VAMPVMMLDAAMIMTAS
Sbjct: 181 GLHGIWLFVGIGFNFIVGVLMTMGLGNYAPELIFFSLMGLSPTVAMPVMMLDAAMIMTAS 240

Query: 250 STQFIKSGRVNWNGFAGLVTGGILGVIVAVLFLTNLDLNSLKTLVVGIVLFTGAMLIRSSF 310
            S+QFIK+ RV+W+GFAG+V+GGI+GV++AV FLTNLD+NSLK LV+ IV FTG MLIRSSF
Sbjct: 241 SSQFIKANRVSWDGFAGIVSGGIIGVLLAVFFLTNLDINSLKLLVIAIVFFTGGMLIRSSF 301
```

A related GBS gene <SEQ ID 8867> and protein <SEQ ID 8868> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      2.32
GvH: Signal Score (-7.5): -5.59
     Possible site: 44
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 8 value:  -8.70 threshold:  0.0
   INTEGRAL    Likelihood = -8.70   Transmembrane  266 - 282 ( 262 - 290)
   INTEGRAL    Likelihood = -7.96   Transmembrane   25 -  41 (  24 -  50)
   INTEGRAL    Likelihood = -6.42   Transmembrane  110 - 126 ( 105 - 140)
   INTEGRAL    Likelihood = -6.26   Transmembrane  194 - 210 ( 190 - 215)
   INTEGRAL    Likelihood = -5.47   Transmembrane  290 - 306 ( 289 - 310)
   INTEGRAL    Likelihood = -4.35   Transmembrane  128 - 144 ( 127 - 147)
   INTEGRAL    Likelihood = -3.29   Transmembrane  157 - 173 ( 156 - 174)
   INTEGRAL    Likelihood = -2.76   Transmembrane  221 - 237 ( 221 - 240)
   PERIPHERAL  Likelihood =  3.87      67
 modified ALOM score:   2.24

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5009> which encodes amino acid sequence <SEQ ID 5010>:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77   Transmembrane  151 - 167 ( 145 - 173)
   INTEGRAL    Likelihood = -9.13   Transmembrane   22 -  38 (  15 -  42)
   INTEGRAL    Likelihood = -8.28   Transmembrane  257 - 273 ( 252 - 281)
   INTEGRAL    Likelihood = -7.48   Transmembrane  281 - 297 ( 275 - 301)
   INTEGRAL    Likelihood = -7.06   Transmembrane  109 - 125 (  95 - 137)
   INTEGRAL    Likelihood = -6.10   Transmembrane  185 - 201 ( 181 - 206)
   INTEGRAL    Likelihood = -1.44   Transmembrane  212 - 228 ( 212 - 229)
   INTEGRAL    Likelihood = -0.27   Transmembrane    5 -  21 (   5 -  21)

----- Final Results -----
              bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  405 bits (1029), Expect = e-115
 Identities = 198/301 (65%), Positives = 228/301 (74%)

Query: 1   LTAKVFFCKLVFMNEMLILRLIQALLVSAMLFIFFMLVKHLKKNKINPFKRFWTGFWIGL 60
            LT  +FFC+L  MNE L+L  IQ +L+ AM+  F+ +V+H++  KINPFKRF+TG WIG
Sbjct: 10  LTVSLFFCRLDIMNETLLLHGIQLILIIAMIITFYQIVRHIRSQKINPFKRFFTGLWIGF 69

Query: 61  LTDALDTLGIGSFATTTTCFKLTKLVTDDRQLPGTMTVAHVLPVLIQSLCFIFVVKVEVX 120
            +TDALDTLGIGSFATTTT FKLTKLV DDR++P TMT AHVLPVL+QSLCFIFVVKVE
Sbjct: 70  VTDALDTLGIGSFATTTTFFKLTKLVEDDRKIPATMTAAHVLPVLLQSLCFIFVVKVEAL 129

Query: 121 XXXXXXXXXFIGAYFGTHITKNWHAPTVQRILGSLLXXXXXXXXXXXXYHPGEHLSDTIH 180
                     FIGA+ G  +TKNWHAPTVQRILG+LL             +PG  +SD++H
Sbjct: 130 TLITMAGAAFIGAFVGAKMTKNWHAPTVQRILGTLLITAAIIMLYRMITNPGAGISDSVH 189

Query: 181 GLHGIWLFVGIGFNFIVGVLMTMGLGNYAPELIFFSLMGLSPTVAMPVMMLDAAMIMTAS 240
            GLHGIWLFVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSP VAMPVMMLDAAMIMTAS
Sbjct: 190 GLHGIWLFVGIGFNFIIGVLMTMGLGNYAPELIFFSLMGLSPAVAMPVMMLDAAMIMTAS 249

Query: 241 SSQFIKANRVSWDXXXXXXXXXXXXXXXXXFFLTNLDINSLKLLVIAIVFFTGGMLIRSSF 301
            S+QFIK+ RV+W+                  FLTNLD+NSLK LV+ IV FTG MLIRSSF
Sbjct: 250 STQFIKSGRVNWNGFAGLVTGGILGVIVAVLFLTNLDLNSLKTLVVGIVLFTGAMLIRSSF 310
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1623

A DNA sequence (GBSx1718) was identified in *S.agalactiae* <SEQ ID 5011> which encodes the amino acid sequence <SEQ ID 5012>. This protein is predicted to be C3-degrading proteinase. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2851(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD37110 GB:AF112358 C3-degrading proteinase [Streptococcus pneumoniae]
```

```
 Identities = 92/240 (38%), Positives = 142/240 (58%), Gaps = 11/240 (4%)

Query: 12  PVLRVNNRDLNIAFYQESLGFKLISEENAIAVFSAWQNKEASFIIEESPTYRTRAVNGTK 71
            P L+ NNR LN  FY E+LG K + EE+A          E   ++EE+P+ RTR V G K
Sbjct: 11  PTLKANNRKLNETFYIETLGMKALLEESAFLSLGDQTGLE-KLVLEEAPSMRTRKVEGRK 69

Query: 72  KLAKIIVKSQDAKDIEKLLANGAQAIQVYQGQNGYAYETVSPEGDLFLLHAEDDLSQLVA 131
            KLA++IVK ++  +IE +L+      ++Y+GQNGYA+E  SPE DL L+HAEDD++ LV
Sbjct: 70  KLARLIVKVENPLEIEGILSKTDSIHRLYKGQNGYAFEIFSPEDDLILIHAEDDIASLVE 129

Query: 132 I-ERPELEKKDDTTGLSNFAFQSISLNVPDAVKAEAFYDKVFAGKFPINLSFKEAQGQDL 190
            + E+PE +    +  LS F   S+ L++P  +  E+F +   + +   +L F  AQGQDL
Sbjct: 130 VGEKPEFQTDLASISLSKFEI-SMELHLPTDI--ESFLE---SSEIGASLDFIPAQGQDL 183

Query: 191 QIAPNETWDIEILECCVNEDTNLNDLKSTFESLGLDVYLDSKEKILVISDTSNIEIWISK 250
             +    TWD+ +L+  VNE  ++  L+  FES   + ++   EK  +  D +N+E+W  +
Sbjct: 184 TVDNTVTWDLSMLKFLVNE-LDIASLRQKFES--TEYFIPKSEKFFLGKDRNNVELWFEE 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5013> which encodes the amino acid sequence <SEQ ID 5014>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3267(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 130/250 (52%), Positives = 177/250 (70%)

Query: 1   MTLFHSLTFKHPVLRVNNRDLNIAFYQESLGFKLISEENAIAVFSAWQNKEASFIIEESP 60
            MTL  ++TFK PVLRVN+RDLNIAFYQ +LG +L+SEENAIA+FS+W   +  F+IEESP
Sbjct: 1   MTLMENITFKTPVLRVNDRDLNIAFYQNNLGLRLVSEENAIAIFSSWGEGQECFVIEESP 60

Query: 61  TYRTRAVNGTKKLAKIIVKSQDAKDIEKLLANGAQAIQVYQGQNGYAYETVSPEGDLFLL 120
            + RTRAV G KK+  I++K+   K+IE+LLA+GA    +++GQNGYA+ET+SPEGD FLL
Sbjct: 61  SVRTRAVEGPKKVNTIVIKTNQPKEIEQLLAHGAHYDALFKGQNGYAFETISPEGDRFLL 120

Query: 121 HAEDDLSQLVAIERPELEKKDDTTGLSNFAFQSISLNVPDAVKAEAFYDKVFAGKFPINL 180
            HAE D+  L   + P LEK     GL+ F F  I LNV    +++AFY  +F+ + PI +
Sbjct: 121 HAEQDIKHLQGTDLPSLEKDATFKGLTQFKFDIIVLNVISEERSKAFYRDLFSDQLPITM 180

Query: 181 SFKEAQGQDLQIAPNETWDIEILECCVNEDTNLNDLKSTFESLGLDVYLDSKEKILVISD 240
             F + +G DL I P+  WD+EILE  V++D ++  LK+T E  G  VY+D K K+LV+SD
Sbjct: 181 DFIQEEGPDLAIDPHIAWDLEILEFQVSKDYDMKVLKATLEEDGHKVYIDKKHKVLVLSD 240

Query: 241 TSNIEIWISK 250
             S IE+W +K
Sbjct: 241 PSQIEVWFTK 250
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1624

A DNA sequence (GBSx1719) was identified in *S.agalactiae* <SEQ ID 5015> which encodes the amino acid sequence <SEQ ID 5016>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                bacterial cytoplasm --- Certainty=0.2510(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC16441 GB:AL450165 putative esterase [Streptomyces coelicolor]
 Identities = 89/323 (27%), Positives = 143/323 (43%), Gaps = 51/323 (15%)

Query: 10  NTVLELIKEQIKDNLYHGASLAIY-ENGEWHEHYLGT-------IDGNEKVKAGLVYDLA 61
            +T+ EL+ E  +  +  GA+ ++    G     + GT       +DG++      V+DLA
Sbjct: 2   STLAELLAEGREQRICSGAAWSVGGPQGPLDRGWTGTRCWDGPPLDGDD------VWDLA 55

Query: 62  SVSKVVGVGTLLAKLVYQGTIDIDKPLRYYYPTFH---HQTLTVRQLATHSSGIDPFIP- 117
            SV+K +  G ++  LV +G + +D  +  Y P +       LTVRQL  H+SGI   +P
Sbjct: 56  SVTKPIA-GLVVMALVERGALGLDDTVGGYLPDYRGGDKAELTVRQLLAHTSGIPGQVPL 114

Query: 118 NRDQLNATQLKDAINHIKVLEDKSFK--YTDINFLLLGFMLEEVLGDSLDKLFKRYIFTP 175
             RD      L +A+  + +      +  Y+   F++LG + E   G+ L+ L +R +  P
Sbjct: 115 YRDHPTRAALLEAVRLLPLTAQPGTRVQYSSQGFIVLGLIAEAAAGEPLEALVERLVCAP 174

Query: 176 FQMKETSFGPRVEAVPTVVGIND---------GIVHDPKAKVLGKHTGSAGLFSTIDDLQ 226
              +++T F P        V   D         G VHD  A VLG   G AGLFST+ D++
Sbjct: 175 LGLRDTVFRPDAGRRARAVATEDCPWRGRRVVGEVHDENAVVLGGVGGHAGLFSTLADME 234

Query: 227 RFSIHYL--------KDDFA-KPLWNNYSLSKSRSLAWD------------IDKDWINHT 265
            R               + FA     +   L+  R+LAW                +   HT
Sbjct: 235 RLGAALAAGGRGLLRPETFALMTAAHTDGLALRRALAWQGRDPVGSPAGEVFGPESYGHT 294

Query: 266 GYTGPFIALNYQKQAAAIFLTNR 288
            G+TG  + ++   +  A+ LTNR
Sbjct: 295 GFTGTSLWVDPATRRYAVLLTNR 317
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3885> which encodes the amino acid sequence <SEQ ID 3886>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.33   Transmembrane   57 -  73 (  57 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 174/302 (57%), Positives = 229/302 (75%), Gaps = 1/302 (0%)

Query: 9    TNTVLELIKEQIKDNLYHGASLAIYENGEWHEHYLGTIDGNEKVKAGLVYDLASVSKVVG 68
            T  V++ I+  +   +Y GASLA++++G W E+++GTIDG   V A LVYDLASVSKVVG
Sbjct: 6    TLAVIKCIENHLHKKVYKGASLALFQSGRWQEYHIGTIDGRRPVDANLVYDLASVSKVVG 65

Query: 69   VGTLLAKLVYQGTIDIDKPLRYYYPTFHHQTLTVRQLATHSSGIDPFIPNRDQLNATQLK 128
            V T+   L+  GT+ +D PL+ YYP+    T+T+RQL TH+SG+DP+IPNRD LNA QL+
Sbjct: 66   VATICNILLNNGTLALDDPLKVYYPSIADATVTIRQLLTHTSGLDPYIPNRDVLNAQQLR 125

Query: 129 DAINHIKVLEDKSFKYTDINFLLLGFMLEEVLGDSLDKLFKRYIFTPFQMKETSFGPRVE 188
             A+NH+   E+K+F YTD+NFLLLGFMLEE+  +SLD++F + IFTPF M  TSFGPR E
Sbjct: 126 KALNHLTQKENKNFYYTDVNFLLLGFMLEELFSESLDQIFDKTIFTPFGMYHTSFGPRPE 185

Query: 189 AVPTVVGINDGIVHDPKAKVLGKHTGSAGLFSTIDDLQRFSIHYLKDDFAKPLWNNYSLS 248
            AVPT+ G++DG VHDPKAK+L KH+GSAGLFST+ DL+ FS HYL D F+  LW NYS
Sbjct: 186 AVPTLKGVSDGEVHDPKAKILKKHSGSAGLFSTLADLESFSNHYLNDPFSDCLWRNYSQQ 245

Query: 249 K-SRSLAWDIDKDWINHTGYTGPFIALNYQKQAAAIFLTNRTFSYDDRPLWIKKRRHVQE 307
               RSL W++D DWI+HTGYTGPF+ LN ++Q AAIFLTNRT+  DD+  W+K+R+ +
```

```
Sbjct: 246 TIERSLGWNLDGDWISHTGYTGPFLMLNKKEQTAAIFLTNRTYDEDDKSKWLKERQLLYN 305

Query: 308 AI 309
           A+
Sbjct: 306 AL 307
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1625

A DNA sequence (GBSx1720) was identified in *S.agalactiae* <SEQ ID 5017> which encodes the amino acid sequence <SEQ ID 5018>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA25177 GB:D21804 FMN-binding protein [Desulfovibrio vulgaris]
 Identities = 53/124 (42%), Positives = 76/124 (60%), Gaps = 2/124 (1%)

Query: 1   MLNHKFLQVLKYEGVVSITSWIELAPHVTNTWNSYLTITDDQRILAPAAGMTHLENDLNN 60
           ML   F +VLK EGVV+I +  E  PH+ NTWNSYL + D  RI+ P  GM   E ++
Sbjct: 1   MLPGTFFEVLKNEGVVAIATQGEDGPHLVNTWNSYLKVLDGNRIVVPVGGMHKTEANVAR 60

Query: 61  NSKIIMTLGSREVEGRDGYQGTGFRIEGTAKLLEAGSDFEIVKEKYPFLRKVLEVTPINV 120
           + +++MTLGSR+V GR+G  GTGF I G+A     G +FE +  ++ + R  L +T ++
Sbjct: 61  DERVLMTLGSRKVAGRNG-PGTGFLIRGSAAFRTDGPEFEAI-ARFKWARAALVITVVSA 118

Query: 121 IQLL 124
            Q L
Sbjct: 119 EQTL 122
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1626

A DNA sequence (GBSx1721) was identified in *S.agalactiae* <SEQ ID 5019> which encodes the amino acid sequence <SEQ ID 5020>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1627

A DNA sequence (GBSx1722) was identified in *S.agalactiae* <SEQ ID 5021> which encodes the amino acid sequence <SEQ ID 5022>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3175(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10123> which encodes amino acid sequence <SEQ ID 10124> was also identified.

The protein has homology to a pyruvate formate-lyase from *S.mutans*:

```
>GP:BAA09085 GB:D50491 Pyruvate formate-lyase [Streptococcus mutans]
 Identities = 709/770 (92%), Positives = 750/770 (97%)

Query: 7    MATVKTNTDIFEQAWEGFKGVDWKEKASIARFVQANYAPYDGDESFLAGATERSLHIKKV 66
            MATVKTNTD+FE+AWEGFKG DWK++ASI+RFVQ NY PYDG ESFLAG TERSLHIKKV
Sbjct: 1    MATVKTNTDVFEKAWEGFKGTDWKDRASISRFVQDNYTPYDGGESFLAGPTERSLHIKKV 60

Query: 67   IEETKAHYEETRFPMDTRVASISELPAGFIDKDNELIFGIQNDELFKLNFMPKGGIRMAE 126
            +EETKAHYEETRFPMDTR+ SI+++PAG+IDK+NELIFGIQNDELFKLNFMPKGGIRMAE
Sbjct: 61   VEETKAHYEETRFPMDTRITSIADIPAGYIDKENELIFGIQNDELFKLNFMPKGGIRMAE 120

Query: 127  TTLKENGYEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSRGRIIG 186
            T LKE+GYEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSRGRIIG
Sbjct: 121  TALKEHGYEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSRGRIIG 180

Query: 187  VYARLAVYGADYLMQEKVNDWNALNDIDEESIRLREEINLQYQALGEVVKLGDLYGVDVR 246
            VYARLA+YGADYLMQEKVNDWN++ +IDEESIRLREEINLQYQALGEVV+LGDLYG+DVR
Sbjct: 181  VYARLALYGADYLMQEKVNDWNSIAEIDEESIRLREEINLQYQALGEVVRLGDLYGLDVR 240

Query: 247  KPAMNTKEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESEIQEFV 306
            KPAMN KEAIQW+NIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESEIQEFV
Sbjct: 241  KPAMNVKEAIQWINIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESEIQEFV 300

Query: 307  DDFVLKLRTVKFARTKAYDALYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTLDNIGN 366
            DDFV+KLRTVKFARTKAYD LYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTLDNIGN
Sbjct: 301  DDFVMKLRTVKFARTKAYDELYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTLDNIGN 360

Query: 367  SPEPNLTVLWSDQLPYAFRRYCMSMSHKHSSIQYEGVSTMAKEGYGEMSCISCCVSPLDP 426
            +PEPNLTVLWS +LPY+FR YCMSMSHKHSSIQYEGV+TMAKEGYGEMSCISCCVSPLDP
Sbjct: 361  APEPNLTVLWSSKLPYSFRHYCMSMSHKHSSIQYEGVTTMAKEGYGEMSCISCCVSPLDP 420

Query: 427  ENEDKRHNLQYFGARVNVMKALLTGLNGGYDDVHKDYKVFDIDPIRDEVLNFDTVKANFE 486
            ENED+RHNLQYFGARVNV+KALLTGLNGGYDDVHKDYKVFD++PIRDEVL+F+TVKANFE
Sbjct: 421  ENEDRRHNLQYFGARVNVLKALLTGLNGGYDDVHKDYKVFDVEPIRDEVLDFETVKANFE 480

Query: 487  KSLDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLPSHVRANMGFGICGFANTVDSLSAIK 546
            K+LDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLP+ V+ANMGFGICGF+NTVDSLSAIK
Sbjct: 481  KALDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLPTRVKANMGFGICGFSNTVDSLSAIK 540

Query: 547  YATVKPIRDEDGYIYDYETVGDFPRYGEDDDRVDSIAEWLLEAFHGRLAKHKLYKDAEAT 606
            YATVKPIRDEDGYIYDYETVG+FPRYGEDDDRVDSIAEWLLEAFH RLA+HKLYKD+EAT
Sbjct: 541  YATVKPIRDEDGYIYDYETVGNFPRYGEDDDRVDSIAEWLLEAFHTRLARHKLYKDSEAT 600

Query: 607  VSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKAKGGWLQNLN 666
```

```
           VSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKA GGWLQNLN
Sbjct: 601 VSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKASGGWLQNLN 660

Query: 667 SLSKLDFAHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLNVMDLKD 726
           SL KLDFAHANDGISLTTQVSP+ALGKTFDEQV NLVT+LDGYFE GGQHVNLNVMDLKD
Sbjct: 661 SLKKLDFAHANDGISLTTQVSPKALGKTFDEQVANLVTILDGYFEGGGQHVNLNVMDLKD 720

Query: 727 VYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDALTN 776
           VYDKIMNGEDVIVRISGYCVNTKYLT EQKTELTQRVFHEVLSMDDA T+
Sbjct: 721 VYDKIMNGEDVIVRISGYCVNTKYLTKEQKTELTQRVFHEVLSMDDAATD 770
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5023> which encodes the amino acid sequence <SEQ ID 5024>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3184(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 701/773 (90%), Positives = 742/773 (95%), Gaps = 1/773 (0%)

Query: 2   FKEKTMATVKTNTDIFEQAWEGFKGVDWKEKASIARFVQANYAPYDGDESFLAGATERSL 61
           FKEK MATVKTNTD+FE+AWEGFKG DWKEKAS++RFVQANY PYDGDESFLAGATERSL
Sbjct: 5   FKEKFMATVKTNTDVFEKAWEGFKGTDWKEKASVSRFVQANYTPYDGDESFLAGATERSL 64

Query: 62  HIKKVIEETKAHYEETRFPMDTRVASISELPAGFIDKDNELIFGIQNDELFKLNFMPKGG 121
           HIKKVIEETKAHYE TRFP DTR  SI+++PAGFIDK+NELI+GIQNDELFKLNFMPKGG
Sbjct: 65  HIKKVIEETKAHYEATRFPYDTRPTSIADIPAGFIDKENELIYGIQNDELFKLNFMPKGG 124

Query: 122 IRMAETTLKENGYEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSR 181
           IRMAETTLKENGYEPDPAVHEIFTKY TTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSR
Sbjct: 125 IRMAETTLKENGYEPDPAVHEIFTKYVTTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSR 184

Query: 182 GRIIGVYARLAVYGADYLMQEKVNDWNALNDIDEESIRLREEINLQYQALGEVVKLGDLY 241
           GRIIGVYARLA+YGADYLMQEKVNDWNA+ +IDEESIRLREE+NLQYQALGEVVKLGDLY
Sbjct: 185 GRIIGVYARLALYGADYLMQEKVNDWNAITEIDEESIRLREEVNLQYQALGEVVKLGDLY 244

Query: 242 GVDVRKPAMNTKEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESE 301
           GVDVR+PA N KEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESE
Sbjct: 245 GVDVRRPAQNVKEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESE 304

Query: 302 IQEFVDDFVLKLRTVKFARTKAYDALYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTL 361
           IQEFVDDFVLKLRTVKF RTKAYDALYSGDPTFITTSMAGMG DGRHRVTKMDYRFLNTL
Sbjct: 305 IQEFVDDFVLKLRTVKFGRTKAYDALYSGDPTFITTSMAGMGNDGRHRVTKMDYRFLNTL 364

Query: 362 DNIGNSPEPNLTVLWSDQLPYAFRRYCMSMSHKHSSIQYEGVSTMAKEGYGEMSCISCCV 421
           DNIGNSPEPNLTVLW+DQLP  FRRYCM MSHKHSSIQYEGV+TMAKEGYGEMSCISCCV
Sbjct: 365 DNIGNSPEPNLTVLWTDQLPETFRRYCMKMSHKHSSIQYEGVTTMAKEGYGEMSCISCCV 424

Query: 422 SPLDPENEDKRHNLQYFGARVNVMKALLTGLNGGYDDVHKDYKVFD-IDPIRDEVLNFDT 480
           SPLDPENE++RHN+QYFGARVNV+KALLTGLNGGYDDVH+DYKVF+ ++PI  EVL +D
Sbjct: 425 SPLDPENEEQRHNIQYFGARVNVLKALLTGLNGGYDDVHRDYKVFNVVEPITSEVLEYDE 484

Query: 481 VKANFEKSLDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLPSHVRANMGFGICGFANTVD 540
           V ANFEKSLDWLTDTYVDA+NIIHYMTDKYNYEAVQMAFLP+H RANMGFGICGFANTVD
Sbjct: 485 VMANFEKSLDWLTDTYVDALNIIHYMTDKYNYEAVQMAFLPTHQRANMGFGICGFANTVD 544

Query: 541 SLSAIKYATVKPIRDEDGYIYDYETVGDFPRYGEDDDRVDSIAEWLLEAFHGRLAKHKLY 600
           +LSAIKYATVK IRDE+GYIYDYE  GDFPRYGEDDDRVD IA+WL+EA+H RLA HKLY
Sbjct: 545 TLSAIKYATVKTIRDENGYIYDYEVTGDFPRYGEDDDRVDDIAKWLMEAYHTRLASHKLY 604

Query: 601 KDAEATVSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKAKGG 660
```

```
           K+AEA+VSLLTITSNVAYSKQTGNSPVH+GV+LNEDG+VN S+VEFFSPGANPSNKAKGG
Sbjct: 605 KNAEASVSLLTITSNVAYSKQTGNSPVHRGVFLNEDGTVNTSQVEFFSPGANPSNKAKGG 664

Query: 661 WLQNLNSLSKLDFAHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLN 720
           WLQNLNSL+KL+F+HANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLN
Sbjct: 665 WLQNLNSLAKLEFSHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLN 724

Query: 721 VMDLKDVYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDA 773
           VMDL DVYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDA
Sbjct: 725 VMDLNDVYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDA 777
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1628

A DNA sequence (GBSx1723) was identified in *S.agalactiae* <SEQ ID 5025> which encodes the amino acid sequence <SEQ ID 5026>. This protein is predicted to be DNA-damage inducible protein P (dinP). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10121> which encodes amino acid sequence <SEQ ID 10122> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF95431 GB:AE004300 DNA-damage-inducible protein P [Vibrio cholerae]
 Identities = 136/349 (38%), Positives = 210/349 (59%), Gaps = 14/349 (4%)

Query: 12  INDTSRKIIHIDMDAFFASVEERDNPSLKGKPVIIGSDPRKTGGRGVVSTCNYEARKFGV 71
           + D  RKIIH+DMD FFA+VE RDNP+ +   + +G   ++   RGV+STCNY+ARKFGV
Sbjct: 1   MQDRIRKIIHVDMDCFFAAVEMRDNPAYREIALAVGGHEKQ---RGVISTCNYQARKFGV 57

Query: 72  HSAMSSKEAYERCPQAIFISGNYQKYRQVGMEVRDIFKKYTDLVEPMSIDEAYLDVTENK 131
            SAM + +A + CPQ   + G    Y+ V  +++ IF++YT L+EP+S+DEAYLDV+E+
Sbjct: 58  RSAMPTAQALKLCPQLHVVPGRMSVYKSVSQQIQTIFQRYTSLIEPLSLDEAYLDVSEST 117

Query: 132 MGIKSAVKLAKMIQYDIWNDVHLTCSAGISYNKFLAKLASDFEKPKGLTLILPDQAQDFL 191
               SA  +A+ I+ DIW +++LT SAG++  KFLAK+ASD  KP GL ++ PD+ Q+ +
Sbjct: 118 AYQGSATLIAQAIRRDIWQELNLTASAGVAPIKFLAKVASDLNKPDGLYVVTPDKVQEMV 177

Query: 192 KPLPIEKFHGVGKRSVEKLHALGVYTGEDLLSLSEISLIDMFGRFGYDLYRKARGINASP 251
             LP+EK  GVGK ++EKLH  G+Y G D+       L+  FGR G  L++K+ GI+
Sbjct: 178 DSLPLEKIPGVGKVALEKLHQAGLYVGADVRRADYRKLLHQFGRLGASLWKKSHGIDERE 237

Query: 252 VKPDRVRKSIGSEKTYGKLLYNEADIKAEISKNVQRVVASLEKNKKVGKTIV---LKVRY 308
           V  +R RKS+G E T+ + +    +    I + +   + +        + I+   +KV++
Sbjct: 238 VVTERERKSVGVEYTFSQNISTFQECWQVIEQKLYPELDARLSRAHPQRGIIKQGIKVKF 297

Query: 309 ADFETLTKRMTLEEYTQDF--QIIDQVAKAIFDTLEESVFGIRLLGVTV 355
           ADF+  T          D+  ++++QV         +    IRLLG++V
Sbjct: 298 ADFQQTTIEHVHPALELDYFHELLEQV------LTRQQGREIRLLGLSV 340
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5027> which encodes the amino acid sequence <SEQ ID 5028>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1921(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 276/363 (76%), Positives = 323/363 (88%)

Query: 6   MLIFPLINDTSRKIIHIDMDAFFASVEERDNPSLKGKPVIIGSDPRKTGGRGVVSTCNYE 65
           MLIFPLINDTSRKIIHIDMDAFFA+VEERDNP+LKGKPV+IG DPR+TGGRGVVSTCNYE
Sbjct: 1   MLIFPLINDTSRKIIHIDMDAFFAAVEERDNPALKGKPVVIGKDPRETGGRGVVSTCNYE 60

Query: 66  ARKFGVHSAMSSKEAYERCPQAIFISGNYQKYRQVGMEVRDIFKKYTDLVEPMSIDEAYL 125
           ARK+G+HSAMSSKEAYERCP+AIFISGNY+KYR VG ++R IFK+YTD+VEPMSIDEAYL
Sbjct: 61  ARKYGIHSAMSSKEAYERCPKAIFISGNYEKYRTVGDQIRRIFKRYTDVVEPMSIDEAYL 120

Query: 126 DVTENKMGIKSAVKLAKMIQYDIWNDVHLTCSAGISYNKFLAKLASDFEKPKGLTLILPD 185
           DVT+NK+GIKSAVK+AK+IQ+DIW +V LTCSAG+SYNKFLAKLASDFEKP GLTL+L +
Sbjct: 121 DVTDNKLGIKSAVKIAKLIQHDIWKEVGLTCSAGVSYNKFLAKLASDFEKPHGLTLVLKE 180

Query: 186 QAQDFLKPLPIEKFHGVGKRSVEKLHALGVYTGEDLLSLSEISLIDMFGRFGYDLYRKAR 245
            A  FL  LPIEKFHGVGK+SV+KLH +G+YTG+DLL++ E++LID FGRFG+DLYRKAR
Sbjct: 181 DALCFLAKLPIEKFHGVGKKSVKKLHDMGIYTGQDLLAVPEMTLIDHFGRFGFDLYRKAR 240

Query: 246 GINASPVKPDRVRKSIGSEKTYGKLLYNEADIKAEISKNVQRVVASLEKNKKVGKTIVLK 305
           GI+ SPVK DR+RKSIGSE+TY KLLY E DIKAEISKNV+RV A L+ +KK+GKTIVLK
Sbjct: 241 GISNSPVKYDRIRKSIGSERTYAKLLYQETDIKAEISKNVKRVAALLQDHKKLGKTIVLK 300

Query: 306 VRYADFETLTKRMTLEEYTQDFQIIDQVAKAIFDTLEESVFGIRLLGVTVTTLENEHEAI 365
           VRYADF TLTKR+TL E T++   I+QVA  IFD+L E+  GIRLLGVT+T LE++   I
Sbjct: 301 VRYADFTTLTKRVTLPELTRNAAQIEQVAGDIFDSLSENPAGIRLLGVTMTNLEDKVADI 360

Query: 366 YLD 368
            LD
Sbjct: 361 SLD 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1629

A DNA sequence (GBSx1724) was identified in *S.agalactiae* <SEQ ID 5029> which encodes the amino acid sequence <SEQ ID 5030>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.11   Transmembrane   70 -  86 (  58 -  92)
   INTEGRAL    Likelihood = -5.20   Transmembrane  105 - 121 ( 100 - 123)
   INTEGRAL    Likelihood = -4.25   Transmembrane  126 - 142 ( 123 - 144)
   INTEGRAL    Likelihood = -2.71   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5031> which encodes the amino acid sequence <SEQ ID 5032>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.00   Transmembrane   69 -  85 (  62 -  93)
   INTEGRAL    Likelihood = -6.85   Transmembrane   16 -  32 (  11 -  37)
   INTEGRAL    Likelihood = -4.30   Transmembrane   99 - 115 (  96 - 121)
   INTEGRAL    Likelihood = -3.66   Transmembrane  126 - 142 ( 121 - 143)

----- Final Results -----
               bacterial membrane --- Certainty=0.6201(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 57/155 (36%), Positives = 96/155 (61%), Gaps = 5/155 (3%)

Query: 1   MVSYEKVRRSLRTATITIIVLNSLSLVFRLFTGISVQLAKTEI-NKGNTGNLPKEHIEAV 59
           M+SYEKVR++L+T+TI II+LN L +V  L     +   ++++ N+     L  E +  +
Sbjct: 1   MISYEKVRQALKTSTIAIIIILNGLGVVLSLMGFAGIFYLQSQLKNEAFRAQLTTEQLAQL 60

Query: 60  LSATTPFMLFVTALIVLVNIAIVIFCIKNLRAIKRNQTVNYLPYYLGFAITVGLVILGFL 119
            S+ TPFM+F++ L VL  IAI++FC +NL  +K+  TV+Y+PY LG  ++V  ++  F
Sbjct: 61  QSSMTPFMIFLSVLNVLAIIAIIVFCAQNLSKLKQGLTVSYIPYILGLILSVIGLVNQFT 120

Query: 120 TTKAPWAIAINIVFQAIFGLLYFHAYQKAQKLNER 154
           TT +     + ++  A++G     A+ KA+ LNE+
Sbjct: 121 TTMSMVGTILILIQAALYGF----AFYKAKTLNEK 151
```

SEQ ID 5030 (GBS227) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 119 (lane 5; MW 21.2kDa).

GBS227-His was purified as shown in Figure 227, lane 8-9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1630

A DNA sequence (GBSx1725) was identified in *S.agalactiae* <SEQ ID 5033> which encodes the amino acid sequence <SEQ ID 5034>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1224(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14706 GB:Z99118 similar to conjugation transfer protein
           [Bacillus subtilis]
 Identities = 328/754 (43%), Positives = 484/754 (63%), Gaps = 25/754 (3%)

Query: 2   EVFFTGTIERIIFENASNFFKILLLEIEDTDSDFDDVEVIITGTMADVIEGEEYTFWGTL 61
           E +  GT+  +I+ N +N + +L +++ +T    +D  V +TG    + E E YTF+G +
Sbjct: 13  EPYLKGTVNTVIYHNDTNLYTVLKVKVTETSEAIEDKAVSVTGYFPALQEEETYTFYGKI 72

Query: 62  TQHPKYGEQLQSVRYERAKPTSG-GLVKYFSSEQFKGIGKKTAQRIVELYGDNTIDKILE 120
             HPK+G Q Q+  +++  PT+  G+++Y SS+ F+GIGKKTA+ IV+  GD+ I+KIL
Sbjct: 73  VTHPKFGLQFQAEHFKKEIPTTKEGIIQYLSSDLFEGIGKKTAEEIVKKLGDSAINKILA 132
```

```
Query: 121 SPEQLSTISGLSKINREAFIAKLKLNYGTEQVLAKLAEYGLSNRAAIQIFDHYKEESLEV 180
               L  +  LSK   +     L+ + G EQ++  L ++G   + +++I+  Y+ E+LE
Sbjct: 133 DASVLYDVPRLSKKKADTLAGALQRHQGLEQIMISLNQFGFGPQLSMKIYQAYESETLEK 192

Query: 181 INENPYQLVEDIQGIGFKIADQLAEQVGIESDSPKRFRAAIIHTLVESSMEQGDTYIEAR 240
           I ENPYQLV+D++GIGF  AD+L  ++G+  + P+R +AAI++TL  + + +G TYIE
Sbjct: 193 IQENPYQLVKDVEGIGFGKADELGSRMGLSGNHPERVKAAILYTLETTCLSEGHTYIETE 252

Query: 241 TLLEKTITLLEEA-----RQIELDPS---IVAKELTNLIAEDKVQHIGTKIFSNTLFFAE 292
            L+  T +LL ++     R  E+D +   I   E  +++ ED       + +  +LF+AE
Sbjct: 253 QLIIDTQSLLNQSAREGQRITEMDAANAIIALGENKDIVIEDG------RCYFPSLFYAE 306

Query: 293 EGIKKNLQRILNQP-LDKQLNHKDIDREIRDIQKSLNIHYDNIQEKAIREALLSKVFILT 351
           + + K ++ I +Q   + Q    +    + ++++ +++ Y   Q++AI++AL S + +LT
Sbjct: 307 QNVAKRVKHIASQTEYENQFPESEFLLALGELEERMDVQYAPSQKEAIQKALSSPMLLLT 366

Query: 352 GGPGTGKTTVINGIIEAYSELHHIDLN----KND--IPIVLAAPTGRAARRMNELTGLPS 405
           GGPGTGKTTVI GI+E Y ELH + L+    K D   PIVLAAPTGRAA+RM+E TGLP+
Sbjct: 367 GGPGTGKTTVIRGIVELYGELHGVSLDPSAYKKDEAFPIVLAAPTGRAAKRMSESTGLPA 426

Query: 406 ATIHRHLGLNGDSDYQSLDDY-LDCSLIIIDEFSMVDTWLANQLFDALDSHTQVIIVGDS 464
            TIHR LG NG   +   +D  ++  L+IIDE SM+D WLAN LF A+  H Q+IIVGD
Sbjct: 427 VTIHRLLGWNGAEGFTHTEDQPIEGKLLIIDEASMLDIWLANHLFKAIPDHIQIIIVGDE 486

Query: 465 DQLPSVGPGQVLADLLNINALPHVKLEKIFRQSEESTIVTLANQMRQGFLPEDFTAKKAD 524
           DQLPSVGPGQVL DLL    +P V+L  I+RQ+E S+IV LA+QM+ G LP + TA   D
Sbjct: 487 DQLPSVGPGQVLRDLLASQVIPTVRLTDIYRQAEGSSIVELAHQMKNGLLPNNLTAPTKD 546

Query: 525 RSYFEASANIIPNMISKIVQSALKSGIEAHEIQILAPMYRGQAGINNLNLIMQNLLNPLK 584
           RS+     + I  ++ K+V +ALK G  A +IQ+LAPMYRG+AGIN LN+++Q++LNP K
Sbjct: 547 RSFIRCGGSQIKEVVEKVVANALKKGYTAKDIQVLAPMYRGKAGINELNVMLQDILNPPK 606

Query: 585 D-NNQFTFNDINFRIGDKVLHLVNDTELNVFNGDIGYITDLIPAKYTESKQDEIYMTFDG 643
           +   +  F D+ +R GDK+L LVN  E NVFNGDIG IT +  AK    K+D   ++FDG
Sbjct: 607 EKRRELKFGDVVYRTGDKILQLVNQPENNVFNGDIGEITSIFYAKENTEKEDMAVVSFDG 666

Query: 644 QEVIYQRKEWLKITLAYAMSIHKSQGSEFQVVILPITRQSGRMLQRNLIYTAITRSKSKL 703
            E+ + +K++ + T AY  SIHKSQGSEF +V+LP+ +   RML+RNL+YTAITR+K  L
Sbjct: 667 NEMTFTKKDFNQFTHAYCCSIHKSQGSEFPIVVLPVVKGYYRMLRRNLLYTAITRAKKFL 726

Query: 704 ILLGEIGAFDFAVKNEGAK-RNTYLIERFENKQE 736
           IL GE  A ++ VKN  A  R T L  R   + E
Sbjct: 727 ILCGEEEALEWGVKNNDATVRQTSLKNRLSVQVE 760
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5035> which encodes the amino acid sequence <SEQ ID 5036>. Analysis of this protein sequence reveals the following:

```
     Possible site: 47

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 232-234
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14706 GB:Z99118 similar to conjugation transfer protein
          [Bacillus subtilis]
 Identities = 318/769 (41%), Positives = 473/769 (61%), Gaps = 29/769 (3%)

Query: 7   GTVDRIIFENQANFFKILLLAIEDTDSDIDDFEIIITGTMADIIEGDDYTFWGELTQHPK 66
           GTV+ +I+ N  N + +L + + +T   I+D  + +TG    + E + YTF+G++  HPK
Sbjct: 18  GTVNTVIYHNDTNLYTVLKVKVTETSEAIEDKAVSVTGYFPALQEEETYTFYGKIVTHPK 77
```

```
Query: 67  YGQQLKLSRYQKIKPSSS-GLVNYFSSDHFKGIGKKTAEKIIALYGHNTIDHILEDPSKL 125
           +G Q +   ++K  P++  G++ Y SSD F+GIGKKTAE+I+   G + I+ IL D S L
Sbjct: 78  FGLQFQAEHFKKEIPTTKEGIIQYLSSDLFEGIGKKTAEEIVKKLGDSAINKILADASVL 137

Query: 126 ETISGLSKANRQAFVAKLKLNYGTEQLIAGLVELGLSNRFALQAFEKYKEEALDLVKENP 185
             +  LSK         L+ + G EQ++  L + G    + +++ ++ Y+ E L+ ++ENP
Sbjct: 138 YDVPRLSKKKADTLAGALQRHQGLEQIMISLNQFGFGPQLSMKIYQAYESETLEKIQENP 197

Query: 186 YQLVEDLQGFGFKMADALAENLGIESDSPKRFRAALLHCLLEESINRGDTYVQARQLLDF 245
           YQLV+D++G GF  AD L   +G+  + P+R +AA+L+ L    ++ G TY++  QL+
Sbjct: 198 YQLVKDVEGIGFGKADELGSRMGLSGNHPERVKAAILYTLETTCLSEGHTYIETEQLIID 257

Query: 246 AITLL-----EDARQVECDPAAVAEQLSE---LIIEGKIKNSDTKLFDASLYFAEEGIAN 297
             +LL     E  R  E D A     L E   ++IE      D + +  SL++AE+ +A
Sbjct: 258 TQSLLNQSAREGQRITEMDAANAIIALGENKDIVIE------DGRCYFPSLFYAEQNVAK 311

Query: 298 NISRLLD-TPLSQSFSHDTIQTTIQAVQKDFAITYDQVQQEAITKALTSKVFLLTGGPGT 356
            +  +   T     F        +  +++   + Y   Q+EAI KAL+S + LLTGGPGT
Sbjct: 312 RVKHIASQTEYENQFPESEFLLALGELEERMDVQYAPSQKEAIQKALSSPMLLLTGGPGT 371

Query: 357 GKTTVIRGILQAYANLHQIDLD----KKD--LPILLAAPTGRAARRMNELTGLPSATIHR 410
           GKTTVIRGI++ Y  LH + LD    KKD   PI+LAAPTGRAA+RM+E TGLP+ TIHR
Sbjct: 372 GKTTVIRGIVELYGELHGVSLDPSAYKKDEAFPIVLAAPTGRAAKRMSESTGLPAVTIHR 431

Query: 411 HLGLNGDNDYQAMEDY-LDCDLLIVDEFSMVDTWLANQLLGAINSTTQVIIVGDSDQLPS 469
            LG NG   +   ED  ++  LLI+DE SM+D WLAN L  AI    Q+IIVGD DQLPS
Sbjct: 432 LLGWNGAEGFTHTEDQPIEGKLLIIDEASMLDIWLANHLFKAIPDHIQIIIVGDEDQLPS 491

Query: 470 VGPGQVLSDLLKVNSLPQIALQKIFRQSQESTIVNLADQMRRGILAADFRDKKADRSYFE 529
           VGPGQVL DLL    +P + L  I+RQ++ S+IV LA QM+ G+L  +      DRS+
Sbjct: 492 VGPGQVLRDLLASQVIPTVRLTDIYRQAEGSSIVELAHQMKNGLLPNNLTAPTKDRSFIR 551

Query: 530 AQAAFIPDMIQKIVLSAIKSGIPAEEIQILAPMYKGQAGINHLNQLMQELLN-PLQGQTE 588
              + I ++++K+V +A+K G  A++IQ+LAPMY+G+AGIN LN ++Q++LN P + + E
Sbjct: 552 CGGSQIKEVVEKVVANALKKGYTAKDIQVLAPMYRGKAGINELNVMLQDILNPPKEKRRE 611

Query: 589 FLFNDTHFRKGDKVLHLVNDAQLNVFNGDIGYITDLIPAKYTESKQDELILDFDGSEVTY 648
             F D  +R GDK+L LVN  + NVFNGDIG IT +  AK    K+D  ++ FDG+E+T+
Sbjct: 612 LKFGDVVYRTGDKILQLVNQPENNVFNGDIGEITSIFYAKENTEKEDMAVVSFDGNEMTF 671

Query: 649 PRNEWLKLTLAYAMSIHKSQGSEFQVVILPITRQSGRLLQRNVIYTAITRSKSKLILLGE 708
            + ++ + T AY  SIHKSQGSEF +V+LP+ +   R+L+RN++YTAITR+K  LIL GE
Sbjct: 672 TKKDFNQFTHAYCCSIHKSQGSEFPIVVLPVVKGYYRMLRRNLLYTAITRAKKFLILCGE 731

Query: 709 YTAFEYAIK-HEGDKRQTYLIERFQEQSDLASSQPNQELKSKEQTSLFS 756
             A E+ +K ++   RQT L  R   Q +    + + EL++ ++   FS
Sbjct: 732 EEALEWGVKNNDATVRQTSLKNRLSVQVE----EMDAELEALQKELPFS 776
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 544/816 (66%), Positives = 665/816 (80%), Gaps = 10/816 (1%)

Query: 1   MEVFFTGTIERIIFENASNFFKILLLEIEDTDSDFDDVEVIITGTMADVIEGEEYTFWGT 60
           ME  FTGT++RIIFEN +NFFKILLL IEDTDSD DD E+IITGTMAD+IEG++YTFWG
Sbjct: 1   MEYVFTGTVDRIIFENQANFFKILLLAIEDTDSDIDDFEIIITGTMADIIEGDDYTFWGE 60

Query: 61  LTQHPKYGEQLQSVRYERAKPTSGGLVKYFSSEQFKGIGKKTAQRIVELYGDNTIDKILE 120
           LTQHPKYG+QL+  RY++ KP+S GLV YFSS+ FKGIGKKTA++I+ LYG NTID ILE
Sbjct: 61  LTQHPKYGQQLKLSRYQKIKPSSSGLVNYFSSDHFKGIGKKTAEKIIALYGHNTIDHILE 120

Query: 121 SPEQLSTISGLSKINREAFIAKLKLNYGTEQVLAKLAEYGLSNRAAIQIFDHYKEESLEV 180
            P +L TISGLSK NR+AF+AKLKLNYGTEQ++A L E GLSNR A+Q F+ YKEE+L++
Sbjct: 121 DPSKLETISGLSKANRQAFVAKLKLNYGTEQLIAGLVELGLSNRFALQAFEKYKEEALDL 180

Query: 181 INENPYQLVEDIQGIGFKIADQLAEQVGIESDSPKRFRAAIIHTLVESSMEQGDTYIEAR 240
           + ENPYQLVED+QG GFK+AD LAE +GIESDSPKRFRAA++H L+E S+ +GDTY++AR
Sbjct: 181 VKENPYQLVEDLQGFGFKMADALAENLGIESDSPKRFRAALLHCLLEESINRGDTYVQAR 240
```

```
Query: 241 TLLEKTITLLEEARQIELDPSIVAKELTNLIAEDKVQHIGTKIFSNTLFFAEEGIKKNLQ 300
            LL+  ITLLE+ARQ+E DP+ VA++L+ LI E K+++  TK+F  +L+FAEEGI  N+
Sbjct: 241 QLLDFAITLLEDARQVECDPAAVAEQLSELIIEGKIKNSDTKLFDASLYFAEEGIANNIS 300

Query: 301 RILNQPLDKQLNHKDIDREIRDIQKSLNIHYDNIQEKAIREALLSKVFILTGGPGTGKTT 360
           R+L+ PL +  +H  I   I+ +QK   I YD +Q++AI +AL SKVF+LTGGPGTGKTT
Sbjct: 301 RLLDTPLSQSFSHDTIQTTIQAVQKDFAITYDQVQQEAITKALTSKVFLLTGGPGTGKTT 360

Query: 361 VINGIIEAYSELHHIDLNKNDIPIVLAAPTGRAARRMNELTGLPSATIHRHLGLNGDSDY 420
           VI GI++AY+ LH IDL+K D+PI+LAAPTGRAARRMNELTGLPSATIHRHLGLNGD+DY
Sbjct: 361 VIRGILQAYANLHQIDLDKKDLPILLAAPTGRAARRMNELTGLPSATIHRHLGLNGDNDY 420

Query: 421 QSLDDYLDCSLIIIDEFSMVDTWLANQLFDALDSHTQVIIVGDSDQLPSVGPGQVLADLL 480
           Q+++DYLDC L+I+DEFSMVDTWLANQL  A++S TQVIIVGDSDQLPSVGPGQVL+DLL
Sbjct: 421 QAMEDYLDCDLLIVDEFSMVDTWLANQLLGAINSTTQVIIVGDSDQLPSVGPGQVLSDLL 480

Query: 481 NINALPHVKLEKIFRQSEESTIVTLANQMRQGFLPEDFTAKKADRSYFEASANIIPNMIS 540
            +N+LP + L+KIFRQS+ESTIV LA+QMR+G L  DF  KKADRSYFEA A  IP+MI
Sbjct: 481 KVNSLPQIALQKIFRQSQESTIVNLADQMRRGILAADFRDKKADRSYFEAQAAFIPDMIQ 540

Query: 541 KIVQSALKSGIEAHEIQILAPMYRGQAGINNLNLIMQNLLNPLKDNNQFTFNDINFRIGD 600
           KIV SA+KSGI A EIQILAPMY+GQAGIN+LN +MQ LLNPL+   +F FND +FR GD
Sbjct: 541 KIVLSAIKSGIPAEEIQILAPMYKGQAGINHLNQLMQELLNPLQGQTEFLFNDTHFRKGD 600

Query: 601 KVLHLVNDTELNVFNGDIGYITDLIPAKYTESKQDEIYMTFDGQEVIYQRKEWLKITLAY 660
           KVLHLVND +LNVFNGDIGYITDLIPAKYTESKQDE+ + FDG EV Y R EWLK+TLAY
Sbjct: 601 KVLHLVNDAQLNVFNGDIGYITDLIPAKYTESKQDELILDFDGSEVTYPRNEWLKLTLAY 660

Query: 661 AMSIHKSQGSEFQVVILPITRQSGRMLQRNLIYTAITRSKSKLILLGEIGAFDFAVKNEG 720
           AMSIHKSQGSEFQVVILPITRQSGR+LQRN+IYTAITRSKSKLILLGE  AF++A+K+EG
Sbjct: 661 AMSIHKSQGSEFQVVILPITRQSGRLLQRNVIYTAITRSKSKLILLGEYTAFEYAIKHEG 720

Query: 721 AKRNTYLIERFENKQEIANSQKIEDSSIDQKI----------DNTIINTSIPKTATPIEQ 770
            KR TYLIERF+ + ++A+SQ  ++    ++           D++  ++S    + P E
Sbjct: 721 DKRQTYLIERFQEQSDLASSQPNQELKSKEQTSLFSNTATLEDDSQKSSSQSTNSNPTEN 780

Query: 771 TNLSKITYRLTEENYLTIDPMIGINQQDISAIFDSK 806
           +      +RLT ENY TID MIG+ + DI+  F  K
Sbjct: 781 SQSDNDDFRLTPENYSTIDSMIGLTESDIALFFQKK 816
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1631

A DNA sequence (GBSx1726) was identified in *S.agalactiae* <SEQ ID 5037> which encodes the amino acid sequence <SEQ ID 5038>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.23    Transmembrane    9 -  25 (   7 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB69116 GB:U90721 signal peptidase I [Streptococcus pneumoniae]
 Identities = 120/201 (59%), Positives = 144/201 (70%), Gaps = 9/201 (4%)

Query: 2   KEFIKEWGVFILILSLFLLSRIFLWQFVKVDGHSMDPTLADKEQLVVLKQTKINRFDIVV 61
           K F+KEWG+F+LILSL  LSRIF W  V+V+GHSMDPTLAD E L V+K   I+RFDIVV
Sbjct: 5   KNFLKEWGLFLLILSLLALSRIFFWSNVRVEGHSMDPTLADGEILFVVKHLPIDRFDIVV 64
```

```
Query: 62  ANEEEGGQKKKIVKRVIGMPGDVIKYKNDTLTINNKKTEEPYLKEYTKLFKKDKLQEKYS 121
           A+EE+G   K IVKRVIGMPGD I+Y+ND L IN+K+T+EPYL +Y K FK DKLQ  YS
Sbjct: 65  AHEEDG--NKDIVKRVIGMPGDTIRYENDKLYINDKETDEPYLADYIKRFKDDKLQSTYS 122

Query: 122 -------YNPLFQDLAQSSTAFTTDSNGSSEFTTVVPKGHYYLVGDDRIVSKDSRAVGPF 174
                      F+ +AQ + AFT D N ++ F+  VP+G Y L+GDDR+VS DSR VG F
Sbjct: 123 GKGFEGNKGTFFRSIAQKAQAFTVDVNYNTNFSFTVPEGEYLLLGDDRLVSSDSRHVGTF 182

Query: 175 KKSTIVGEVKFRFWPIRRFGT 195
           K   I GE KFRFWPI R GT
Sbjct: 183 KAKDITGEAKFRFWPITRIGT 203
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5039> which encodes the amino acid sequence <SEQ ID 5040>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.50   Transmembrane   35 -  51 (  35 -  51)

----- Final Results -----
                 bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9157> which encodes the amino acid sequence <SEQ ID 9158>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                  bacterial outside --- Certainty= 0.300(Affirmative) < succ>
                 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/197 (66%), Positives = 152/197 (76%)

Query: 1   MKEFIKEWGVFILILSLFLLSRIFLWQFVKVDGHSMDPTLADKEQLVVLKQTKINRFDIV 60
           MK+FIKEWG F L L LF LSR+FLWQ VKVDGHSMDPTLA  E+L+V  Q +I+RFDIV
Sbjct: 23  MKQFIKEWGPFTLFLILFGLSRLFLWQAVKVDGHSMDPTLAHGERLIVFNQARIDRFDIV 82

Query: 61  VANEEEGGQKKKIVKRVIGMPGDVIKYKNDTLTINNKKTEEPYLKEYTKLFKKDKLQEKY 120
           VA EEE GQKK+IVKRVIG+PGD I Y +DTL IN KKT EPYL EY K FK DKLQ+ Y
Sbjct: 83  VAQEEENGQKKEIVKRVIGLPGDTISYNDDTLYINGKKTVEPYLAEYLKQFKNDKLQKTY 142

Query: 121 SYNPLFQDLAQSSTAFTTDSNGSSEFTTVVPKGHYYLVGDDRIVSKDSRAVGPFKKSTIV 180
           +YN LFQ LA++S AFTT+S G + F   VPKG Y L+GDDRIVS+DSR VG FKK  ++
Sbjct: 143 AYNTLFQQLAETSDAFTTNSEGQTRFEMSVPKGEYLLLGDDRIVSRDSREVGSFKKENLI 202

Query: 181 GEVKFRFWPIRRFGTIN 197
           GEVK RFWP+ +    N
Sbjct: 203 GEVKARFWPLNKMTVFN 219
```

SEQ ID 5038 (GBS268) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 4; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 9; MW 25.3kDa) and in Figure 160 (lane 2-4; MW 25.3kDa).

GBS268-His was purified as shown in Figure 222, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1632

A DNA sequence (GBSx1727) was identified in *S.agalactiae* <SEQ ID 5041> which encodes the amino acid sequence <SEQ ID 5042>. This protein is predicted to be ribonuclease HIII (rnhB). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4728(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10119> which encodes amino acid sequence <SEQ ID 10120> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45437 GB:U93576 ribonuclease HII [Streptococcus pneumoniae]
 Identities = 176/282 (62%), Positives = 219/282 (77%), Gaps = 13/282 (4%)

Query: 16  EKIRTDLAQHHISNNNPYVVFSAKISGATVLLYTSGKLVFQGSNASHIAQKYGF--IEQK 73
           E  +T LA     + NPY+ +  K+  ATV +YTSGK++ QG  A   A  +G+  +EQ
Sbjct: 18  EHYQTSLAP----SKNPYIRYFLKLPQATVSIYTSGKILLQGEGAEKYASFFGYQAVEQ- 72

Query: 74  ESCSSESQDIPIIGTDEVGNGSYFGGLAVVASFVTPKDHAYLKKLGVGDSKTLTDQKIKQ 133
                +  Q++P+IGTDEVGNGSYFGGLAVVA+FVTP  H +L+KLGVGDSKTLTDQKI+Q
Sbjct: 73  ----TSGQNLPLIGTDEVGNGSYFGGLAVVAAFVTPDQHDFLRKLGVGDSKTLTDQKIRQ 128

Query: 134 IAPLLEKAIPHKALLLSPQKYNQVVSPNNKHNAVSVKVALHNQAIFLLLQDGFEPEKIVI 193
           IAP+L++ I H+ALLLSP KYN+V+   +++NAVSVKVALHNQAI+LLLQ G +PEKIVI
Sbjct: 129 IAPILKEKIQHQALLLSPSKYNEVIG--DRYNAVSVKVALHNQAIYLLLQKGVQPEKIVI 186

Query: 194 DAFTSSKNYQNYLKNEKNQFKQTITLEEKAENKYLAVAVSSIIARNLFLENLNKLSDDVG 253
           DAFTS+KNY  YL  E N+F   I+LEEKAE KYLAVAVSS+IAR+LFLENL  L  ++G
Sbjct: 187 DAFTSAKNYDKYLAQETNRFSNPISLEEKAEGKYLAVAVSSVIARDLFLENLENLGRELG 246

Query: 254 YKLPSGAGHQSDKVASQLLKAYGISSLEHCAKLHFANTKKAQ 295
           Y+LPSGAG  SDKVASQ+L+AYG+  L  CAKLHF NT+KA+
Sbjct: 247 YQLPSGAGTASDKVASQILQAYGMQGLNFCAKLHFKNTEKAK 288
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5043> which encodes the amino acid sequence <SEQ ID 5044>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2148(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 194/298 (65%), Positives = 240/298 (80%), Gaps = 2/298 (0%)

Query: 3   MNTIVMQADKKLQEKIRTDLAQHHISNNNPYVVFSAKISGATVLLYTSGKLVFQGSNASH 62
           MNT+V++ D  L + ++  LA + IS+ N YV F+AK +G TVLLY SGKLV QG+ A+
Sbjct: 1   MNTLVLKIDAILSKHLKKQLAPYTISSQNTYVAFAAKKNGVTVLLYKSGKLVLQGNGANA 60
```

```
Query: 63  IAQKYGFIEQKE--SCSSESQDIPIIGTDEVGNGSYFGGLAVVASFVTPKDHAYLKKLGV 120
           +AQ+      K     S+ SQDIPIIG+DEVGNGSYFGG+AVVASFV PKDH++LKKLGV
Sbjct: 61  LAQELNLPVAKTVFEASNNSQDIPIIGSDEVGNGSYFGGIAVVASFVDPKDHSFLKKLGV 120

Query: 121 GDSKTLTDQKIKQIAPLLEKAIPHKALLLSPQKYNQVVSPNNKHNAVSVKVALHNQAIFL 180
            DSK L+D+ I+QIAPLLEK IPH++LLLSP+KYN++V  +  +NA+S+KVALHNQAIFL
Sbjct: 121 DDSKKLSDKTIQQIAPLLEKQIPHQSLLLSPKKYNELVGKSKPYNAISIKVALHNQAIFL 180

Query: 181 LLQDGFEPEKIVIDAFTSSKNYQNYLKNEKNQFKQTITLEEKAENKYLAVAVSSIIARNL 240
           LLQ G +P++IVIDAFTS  NY+ +LK EKN F   +T +EKAE+ YLAVAVSSIIARNL
Sbjct: 181 LLQKGIQPKQIVIDAFTSQSNYEKHLKKEKNHFPNPLTFQEKAESHYLAVAVSSIIARNL 240

Query: 241 FLENLNKLSDDVGYKLPSGAGHQSDKVASQLLKAYGISSLEHCAKLHFANTKKAQALL 298
           FL+NL++L  D+GY+LPSGAG  SDKVASQLL AYG+SSLE+ AKLHFANT KAQALL
Sbjct: 241 FLDNLDQLGQDLGYQLPSGAGSASDKVASQLLAAYGMSSLEYSAKLHFANTHKAQALL 298
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1633

A DNA sequence (GBSx1728) was identified in *S.agalactiae* <SEQ ID 5045> which encodes the amino acid sequence <SEQ ID 5046>. This protein is predicted to be heat shock protein 70. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3874(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5047> which encodes the amino acid sequence <SEQ ID 5048>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3442(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 65/92 (70%), Positives = 76/92 (81%)

Query: 11  NRYKFVFGDKPLTLTTDKDNLFMEEIERVATEKYEAIKEKLPNADNETIAILMAINALSV 70
           NRYKF FG+K LTLTTDKDNLFMEE+ERVA EKY+A+K  LP AD+ETIAILMAIN LS
Sbjct: 5   NRYKFTFGEKTLTLTTDKDNLFMEEVERVAKEKYQALKNHLPEADDETIAILMAINTLST 64

Query: 71  QLSREIDIEKMEDELNKLRSKTISDIKEKVSE 102
           QLSREI IEKME E+  LR KT+  ++EK ++
Sbjct: 65  QLSREIAIEKMEAEILDLRQKTLVGLQEKANQ 96
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1634

A DNA sequence (GBSx1729) was identified in *S.agalactiae* <SEQ ID 5049> which encodes the amino acid sequence <SEQ ID 5050>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.99   Transmembrane  124 - 140 ( 114 - 148)
   INTEGRAL    Likelihood = -5.84   Transmembrane   22 -  38 (  21 -  40)
   INTEGRAL    Likelihood = -4.88   Transmembrane    2 -  18 (   1 -  20)
   INTEGRAL    Likelihood = -1.97   Transmembrane   84 - 100 (  84 - 100)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06827 GB:AP001517 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 59/182 (32%), Positives = 98/182 (53%), Gaps = 14/182 (7%)

Query: 1   MLSLLLLIIVIWHFYIGYSRGIFLQVFYVLMSMVSLMIASQFYQELASQITLWVPYS--N 58
           MLS++LL I++  F+IG  RG+ LQ+ ++L  + +  +A ++Y  +A+ I LW+PY   +
Sbjct: 1   MLSVILLFILLCSFFIGKRRGLILQLVHLLGFVAAFFVAYKYYAPVATYIRLWIPYPQFS 60

Query: 59  PVQGVEVYFFKDISKFQLSHVYYAGVAFVFIY----SLSYLVGRLLGVLLHLAPVEHFDS 114
           P   V +     I  F   +VYY+G+AF ++      L ++VG +L  L HL     +
Sbjct: 61  PDSPVTML----IEAFNFENVYYSGIAFALLFIGTKILLHIVGSMLDFLTHLPILRSV-- 114

Query: 115 LQNNIISGFLAVLVCLLFMSMCLTILATVPMSFVQEKLWNSLFVRFLINDLPFFSQFLVR 174
             N  +  G L  +   L M + L + A +P+  VQ  L  SL  +F++N  PF S+F+
Sbjct: 115 --NGWLGGILGFVEVYLIMFVLLYVGALLPIETVQTHLNQSLVAQFIMNHTPFLSEFIRN 172

Query: 175 TW 176
            W
Sbjct: 173 LW 174
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5051> which encodes the amino acid sequence <SEQ ID 5052>. Analysis of this protein sequence reveals the following:

```
     Possible site: 59
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.17   Transmembrane  124 - 140 ( 117 - 148)
   INTEGRAL    Likelihood = -4.73   Transmembrane   84 - 100 (  78 - 105)
   INTEGRAL    Likelihood = -0.00   Transmembrane  156 - 172 ( 156 - 172)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06827 GB:AP001517 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 57/177 (32%), Positives = 98/177 (55%), Gaps = 2/177 (1%)

Query: 1   MLSLLIVLILTWNFYIGYSRGIILQSFYVLGALLSLLVANRFYIGLAHKLTLWIPYSNPV 60
           MLS++++ IL  +F+IG  RG+ILQ  ++LG + +  VA ++Y  +A  + LWIPY
Sbjct: 1   MLSVILLFILLCSFFIGKRRGLILQLVHLLGFVAAFFVAYKYYAPVATYIRLWIPYPQFS 60

Query: 61  EGTSVFFFKSVDIFVLDKVYYAGLAFFIIFLLGYALSRFLGIFVHFLLLNYFDNQWTKCL 120
             + V     ++ F  + VYY+G+AF ++F+    L   +G  + FL            L
Sbjct: 61  PDSPVTML--IEAFNFENVYYSGIAFALLFIGTKILLHIVGSMLDFLTHLPILRSVNGWL 118
```

```
Query: 121 SGGLAFLVSLLFLNMLLSIFATVPMPFLQHYLHSSFLARLVIEHLPPLTIIIQKLWI 177
            G L F+   L + +LL + A +P+  +Q +L+ S +A+ ++ H P L+  I+ LWI
Sbjct: 119 GGILGFVEVYLIMFVLLYVGALLPIETVQTHLNQSLVAQFIMNHTPFLSEFIRNLWI 175
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/176 (49%), Positives = 123/176 (69%)

Query: 1   MLSLLLLIIVIWHFYIGYSRGIFLQVFYVLMSMVSLMIASQFYQELASQITLWVPYSNPV 60
           MLSLL+++I+ W+FYIGYSRGI LQ FYVL +++SL++A++FY  LA ++TLW+PYSNPV
Sbjct: 1   MLSLLIVLILTWNFYIGYSRGIILQSFYVLGALLSLLVANRFYIGLAHKLTLWIPYSNPV 60

Query: 61  QGVEVYFFKDISKFQLSHVYYAGVAFVFIYSLSYLVGRLLGVLLHLAPVEHFDSLQNNII 120
           +G  V+FFK +  F L  VYYAG+AF  I+ L Y + R LG+ +H   + +FD+     +
Sbjct: 61  EGTSVFFFKSVDIFVLDKVYYAGLAFFIIFLLGYALSRFLGIFVHFLLLNYFDNQWTKCL 120

Query: 121 SGFLAVLVCLLFMSMCLTILATVPMSFVQEKLWNSLFVRFLINDLPFFSQFLVRTW 176
           SG LA LV LLF++M L+I ATVPM F+Q  L +S   R +I  LP  +  + + W
Sbjct: 121 SGGLAFLVSLLFLNMLLSIFATVPMPFLQHYLHSSFLARLVIEHLPPLTIIIQKLW 176
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1635

A DNA sequence (GBSx1730) was identified in *S.agalactiae* <SEQ ID 5053> which encodes the amino acid sequence <SEQ ID 5054>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4176(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10117> which encodes amino acid sequence <SEQ ID 10118> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14818 GB:Z99118 similar to DNA mismatch repair protein
          [Bacillus subtilis]
 Identities = 320/790 (40%), Positives = 466/790 (58%), Gaps = 18/790 (2%)

Query: 10  MNNKILEQLEFNKVKELILPYLKTEQSQEELSELEPMTEAPKIEKSFNEISDMEQIFVEH 69
           M  K+L  LEF+KVKE ++ +  +   +E L EL+P     +I+K  +E+ +   I
Sbjct: 1   MQQKVLSALEFHKVKEQVIGHAASSLGKEMLLELKPSASIDEIKKQLDEVDEASDIIRLR 60

Query: 70  HSFGIVSLSSISESLKRLELSADLNIQELLAIKKVLQSSSDMIHFYSDL--DNVSFQSLD 127
                  L  I  +L+R E+ + L+  E   I  +L +    M HF + +  D V    +
Sbjct: 61  GQAPFGGLVDIRGALRRAEIGSVLSPSEFTEISGLLYAVKQMKHFITQMAEDGVDIPLIH 120

Query: 128 RLFENLEQFPNLQGSFQA-INDGGFLEHFASPELERIRRQLTNSERRVRQILQDMLKEKA 186
           +  E L    +L+    + I+D G +   AS  L  IR QL   E RVR  L+ ML+  +
Sbjct: 121 QHAEQLITLSDLERDINSCIDDHGEVLDHASETLRGIRTQLRTLESRVRDRLESMLRSSS 180

Query: 187 --ELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNEEITQL 244
             ++LS+ ++  R+ R V+PVK  YR+   G+VHD SSSG+T++IEP+A+V +N  + Q
Sbjct: 181 ASKMLSDTIVTIRNDRFVIPVKQEYRSSYGGIVHDTSSSGATLFIEPQAIVDMNNSLQQA 240

Query: 245 RADERHEESRILHAFSDLLRPHVATIRNNAWILGHLDFVRAKYLFMSDNKATIPEISNDS 304
           +  E+ E  RIL   ++    +   +  +  +L  LDF+ AK  +    KAT P +++
Sbjct: 241 KVKEKQEIERILRVLTEKTAEYTEELFLDLQVLQTLDFIFAKARYAKAVKATKPIMNDTG 300
```

```
Query: 305 TLALINVRHPLL--SNPVANDLHFDQDLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGLP 362
            + L   RHPLL     VAND+   +D + IVITGPNTGGKT+ LKTLGL  LM QSGL
Sbjct: 301 FIRLKKARHPLLPPDQVVANDIELGRDFSTIVITGPNTGGKTVTLKTLGLLTLMAQSGLH 360

Query: 363 VLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGAG 422
           + AD+GS+ AVF ++FADIGDEQSIEQSLSTFSSHM +IV IL + + NSLVLFDELGAG
Sbjct: 361 IPADEGSEAAVFEHVFADIGDEQSIEQSLSTFSSHMVNIVGILEQVNENSLVLFDELGAG 420

Query: 423 TDPQEGASLAMAILEHLRLSNIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTYR 482
           TDPQEGA+LAM+IL+ +  +N + +ATTHYPELKAYG     V NAS+EFD ETLSPTY+
Sbjct: 421 TDPQEGAALAMSILDDVHRTNARVLATTHYPELKAYGYNREGVMNASVEFDIETLSPTYK 480

Query: 483 FMQGVPGRSNAFEIASRLGLAPFIVKQAK-QMTDSDSDVNRIIEQLEAQTLETRRRLDHI 541
            + GVPGRSNAFEI+ RLGL   I+ QAK +MT   ++V+ +I  LE         L
Sbjct: 481 LLIGVPGRSNAFEISKRLGLPDHIIGQAKSEMTAEHNEVDTMIASLEQSKKRAEEELSET 540

Query: 542 KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKL----ND 597
           + + +E  K ++ +++   E + ++DK LE+  Q+A E V  A+ E++ I+ +L     +
Sbjct: 541 ESIRKEAEKLHKELQQQIIELNSKKDKMLEEAEQQAAEKVKAAMKEAEDIIHELRTIKEE 600

Query: 598 KSQLKPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTL 657
               K HE+I+AK +++   P  + SK     K +K    R  + GD++ V ++GQ+GTL
Sbjct: 601 HKSFKDHELINAKKRLEGAMPAFEKSKKPEKPKTQK----RDFKPGDEVKVLTFGQKGTL 656

Query: 658 TSQLKDGRWEAQVGIIKMTLTQDEFTLVRVQEEQKVKSKQINVVKKADSSGPRARLDLRG 717
             +     W  Q+GI+KM + + +   ++    E K K K I  VK  D       LDLRG
Sbjct: 657 LEKTGGNEWNVQIGILKMKVKEKDLEFIKSAPEPK-KEKMITAVKGKDYH-VSLELDLRG 714

Query: 718 KRYEEAMQELDNFIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNA 777
           +RYE A+  ++ ++D A+L    +V IIHG GTG +R+GV   L+ ++ VK   +
Sbjct: 715 ERYENALSRVEKYLDDAVLAGYPRVSIIHGKGTGALRKGVQDLLKNHRSVKSSRFGEAGE 774

Query: 778 GGSGATIVTL 787
           GGSG T+V L
Sbjct: 775 GGSGVTVVEL 784
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5055> which encodes the amino acid sequence <SEQ ID 5056>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3843(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 775/787 (98%), Positives = 781/787 (98%)

Query: 2   INLGIMKSMNNKILEQLEFNKVKELILPYLKTEQSQEELSELEPMTEAPKIEKSFNEISD 61
            I LGIMKSMNNKILEQLEFNKVKEL+LPYLKTEQSQEEL ELEPMTEAPKIEKSFNEISD
Sbjct: 32  IILGIMKSMNNKILEQLEFNKVKELLLPYLKTEQSQEELLELEPMTEAPKIEKSFNEISD 91

Query: 62  MEQIFVEHHSFGIVSLSSISESLKRLELSADLNIQELLAIKKVLQSSSDMIHFYSDLDNV 121
            MEQIFVEHHSFGIVSLSSISESLKRLELS DLNIQELLAIKKVLQSSSDMIHFYSDLDNV
Sbjct: 92  MEQIFVEHHSFGIVSLSSISESLKRLELSTDLNIQELLAIKKVLQSSSDMIHFYSDLDNV 151

Query: 122 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQILQDM 181
            SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQILQDM
Sbjct: 152 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQILQDM 211

Query: 182 LKEKAELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNEEI 241
            LKEKAELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNEEI
Sbjct: 212 LKEKAELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNEEI 271
```

```
Query: 242 TQLRADERHEESRILHAFSDLLRPHVATIRNNAWILGHLDFVRAKYLFMSDNKATIPEIS 301
           TQLRADERHEE RILHAFSDLLRPHVATIRNNAWILGHLDFVRAKYLFMSDNKATIP+IS
Sbjct: 272 TQLRADERHEEGRILHAFSDLLRPHVATIRNNAWILGHLDFVRAKYLFMSDNKATIPKIS 331

Query: 302 NDSTLALINVRHPLLSNPVANDLHFDQDLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGL 361
           NDSTLALINVRHPLLSNPVANDLHFD DLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGL
Sbjct: 332 NDSTLALINVRHPLLSNPVANDLHFDHDLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGL 391

Query: 362 PVLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGA 421
           PVLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGA
Sbjct: 392 PVLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGA 451

Query: 422 GTDPQEGASLAMAILEHLRLSNIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTY 481
           GTDPQEGASLAMAILEHLRLS+IKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTY
Sbjct: 452 GTDPQEGASLAMAILEHLRLSHIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTY 511

Query: 482 RFMQGVPGRSNAFEIASRLGLAPFIVKQAKQMTDSDSDVNRIIEQLEAQTLETRRRLDHI 541
           RFMQGVPGRSNAFEIASRLGLAPFIVKQAKQMTDSDSDVNRIIEQLEAQTLETRRRLDHI
Sbjct: 512 RFMQGVPGRSNAFEIASRLGLAPFIVKQAKQMTDSDSDVNRIIEQLEAQTLETRRRLDHI 571

Query: 542 KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL 601
           KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL
Sbjct: 572 KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL 631

Query: 602 KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTLTSQL 661
           KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTLTSQL
Sbjct: 632 KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTLTSQL 691

Query: 662 KDGRWEAQVGIIKMTLTQDEFTLVRVQEEQKVKSKQINVVKKADSSGPRARLDLRGKRYE 721
           KDGRWEAQVGIIKMTLTQDEF+LVRVQEEQKVK+KQINVVKKAD SGPRARLDLRGKRYE
Sbjct: 692 KDGRWEAQVGIIKMTLTQDEFSLVRVQEEQKVKNKQINVVKKADGSGPRARLDLRGKRYE 751

Query: 722 EAMQELDNFIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNAGGSG 781
           EAMQELD+FIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNAGGSG
Sbjct: 752 EAMQELDHFIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNAGGSG 811

Query: 782 ATIVTLG 788
           ATIVTLG
Sbjct: 812 ATIVTLG 818
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1636

A DNA sequence (GBSx1731) was identified in *S.agalactiae* <SEQ ID 5057> which encodes the amino acid sequence <SEQ ID 5058>. This protein is predicted to be thioredoxin (trxA). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10115> which encodes amino acid sequence <SEQ ID 10116> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB40815 GB:AJ133006 thioredoxin [Listeria monocytogenes] (ver
```

```
          2)
 Identities = 64/100 (64%), Positives = 78/100 (78%), Gaps = 1/100 (1%)

Query: 15  MALEVTDATFVEETKEGLVLIDFWATWCGPCRMQAPILEQLSQEIDEDELKILKMDVDEN 74
            M  E+TDATF +ET EGLVL DFWATWCGPCRM AP+LE++ +E  E  LKI+KMDVDEN
Sbjct: 1   MVKEITDATFEQETSEGLVLTDFWATWCGPCRMVAPVLEEIQEERGE-ALKIVKMDVDEN 59

Query: 75  PETARQFGIMSIPTLMFKKDGEVVKQVAGVHTKDQLKAII 114
            PET   FG+MSIPTL+ KKDGEVV+ + G   K++L  +I
Sbjct: 60  PETPGSFGVMSIPTLLIKKDGEVVETIIGYRPKEELDEVI 99
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5059> which encodes the amino acid sequence <SEQ ID 5060>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1637

A DNA sequence (GBSx1732) was identified in *S.agalactiae* <SEQ ID 5061> which encodes the amino acid sequence <SEQ ID 5062>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.54   Transmembrane  170 - 186 ( 167 - 191)
   INTEGRAL    Likelihood = -5.52   Transmembrane   87 - 103 (  86 - 107)
   INTEGRAL    Likelihood = -4.62   Transmembrane  105 - 121 ( 104 - 126)

----- Final Results -----
               bacterial membrane --- Certainty=0.4015(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60798 GB:X87369 ORF3 [Clostridium perfringens]
 Identities = 27/67 (40%), Positives = 52/67 (77%)

Query: 1  MEIGQQIIRYRKQQALSQEELAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ 60
           M++ +++   RK++ LSQE+LAEK+ +SRQ++S WE+ ++ PD++ L++LS+++ V++D
Sbjct: 1  MKLAEKLQLMRKREGLSQEDLAEKLGISRQAVSKWESGQSVPDLNKLIILSELYNVTIDY 60

Query: 61 LIKGDIE 67
           L+K   E
Sbjct: 61 LVKETYE 67
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1739> which encodes the amino acid sequence <SEQ ID 1740>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.86   Transmembrane  173 - 189 ( 169 - 194)
   INTEGRAL    Likelihood = -5.52   Transmembrane   90 - 106 (  89 - 110)
   INTEGRAL    Likelihood = -4.62   Transmembrane  108 - 124 ( 107 - 129)

----- Final Results -----
```

```
                  bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 187/195 (95%), Positives = 191/195 (97%)

Query: 1   MEIGQQIIRYRKQQALSQEELAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ 60
           MEIGQQIIRYRKQQALSQE+LAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ
Sbjct: 4   MEIGQQIIRYRKQQALSQEKLAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ 63

Query: 61  LIKGDIEKMKYTITQVDKKNFERDTKVMVTLMILLMISSYPLVYFLEWLGLGIFVLLSII 120
           LIKGDIEKMKYTITQVDKKNF+RDTKVMVTLMILLMISSYPLVYFLEWLGLGIFVLLSII
Sbjct: 64  LIKGDIEKMKYTITQVDKKNFKRDTKVMVTLMILLMISSYPLVYFLEWLGLGIFVLLSII 123

Query: 121 TMTYANRVERFKKKYDVQTYKEILAVSSGKLLDEIEKREERAKLPYQKPLIVTVFFLITV 180
           TMTYANRVERFKKKYDVQ YKEILAVS+GKLLDEIEKREERA LPYQKPLIVTVFFLITV
Sbjct: 124 TMTYANRVERFKKKYDVQPYKEILAVSNGKLLDEIEKREERATLPYQKPLIVTVFFLITV 183

Query: 181 ATFFASRFIFTWLFH 195
           A  FASRF+FTWLFH
Sbjct: 184 AFAFASRFMFTWLFH 198
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1638

A DNA sequence (GBSx1733) was identified in *S.agalactiae* <SEQ ID 5063> which encodes the amino acid sequence <SEQ ID 5064>. This protein is predicted to be adenine glycosylase (mutY). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2385(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9425> which encodes amino acid sequence <SEQ ID 9426> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04650 GB:AP001510 adenine glycosylase [Bacillus halodurans]
 Identities = 130/331 (39%), Positives = 190/331 (57%), Gaps = 15/331 (4%)

Query: 1   MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV 60
           MLQQT+V+TVIPYY+ F+  FP ++ LA A E+Q+LKAWEGLGYYSR RN+Q A ++V+
Sbjct: 45  MLQQTRVDTVIPYYQAFMRQFPTLETLAYAEEDQVLKAWEGLGYYSRARNLQSAVREVVE 104

Query: 61  DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 120
            +GG  P T  +I+ LKG+GPYTAGAI SI+++ PEPAVDGNVMRV++R+  +  DI
Sbjct: 105 SYGGEVPSTRKEISKLKGVGPYTAGAILSIAYDQPEPAVDGNVMRVLSRVLYIEEDIAKV 164

Query: 121 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYSK 180
           K R +F++++  LI  + P  FNQ LM+LG  + +  +P     P+R    A+  G   +
Sbjct: 165 KTRTLFESLLYDLISKENPSFFNQGLMELGALVCTPTSPGCLLCPVRDHCRAFAAGVQEQ 224

Query: 181 YPIKNTKKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD 240
            PIK  KKKPK  ++ A VIRN+ GQ L+E+  +  LL   W FP +E         L
Sbjct: 225 LPIKAKKKKPKAKQLIAAVIRNEKGQVLIERRPEKGLLAKLWQFPNVE---------LES 275
```

```
Query: 241 DNQSNPIIWQTQNETFQREYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKAT-DLPN 299
              +  ++    +E F  +  +       +   ++H FSH  W I + E  VK    L +
Sbjct: 276 TKNAQQVLGDYIHERFHLDAAV-----GEYVQTVEHVFSHLIWNIRVYEATVKGVPSLND 330

Query: 300 APHLKWVAIEDFSLYPFATPQKKMLETYLKQ 330
                WV       Y F    +K+++  L++
Sbjct: 331 KYEADWVDDRTIENYAFPVSHQKIIQGNLRK 361
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5065> which encodes the amino acid sequence <SEQ ID 5066>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3579(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 330/333 (99%), Positives = 331/333 (99%)

Query: 1   MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV 60
           MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV
Sbjct: 52  MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV 111

Query: 61  DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 120
           DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP
Sbjct: 112 DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 171

Query: 121 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYSK 180
           KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTY K
Sbjct: 172 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYGK 231

Query: 181 YPIKNTKKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD 240
           YPIKN KKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD
Sbjct: 232 YPIKNPKKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD 291

Query: 241 DNQSNPIIWQTQNETFQREYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKATDLPNA 300
           DNQSNPIIWQTQNETF+REYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKATDLPNA
Sbjct: 292 DNQSNPIIWQTQNETFEREYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKATDLPNA 351

Query: 301 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA 333
           PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA
Sbjct: 352 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA 384
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1639

A DNA sequence (GBSx1734) was identified in *S.agalactiae* <SEQ ID 5067> which encodes the amino acid sequence <SEQ ID 5068>. This protein is predicted to be maltose/maltodextrin transport system (malG). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.30   Transmembrane   14 -  30 (   5 -  35)
   INTEGRAL    Likelihood = -6.95   Transmembrane  248 - 264 ( 242 - 267)
   INTEGRAL    Likelihood = -5.15   Transmembrane   75 -  91 (  74 -  94)
   INTEGRAL    Likelihood = -3.19   Transmembrane  110 - 126 ( 110 - 127)
   INTEGRAL    Likelihood = -2.13   Transmembrane  141 - 157 ( 138 - 157)
```

```
     INTEGRAL    Likelihood = -0.32   Transmembrane  188 - 204 ( 188 - 204)

----- Final Results -----
                 bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06643 GB:AP001517 maltose/maltodextrin transport system
            (permease) [Bacillus halodurans]
 Identities = 117/281 (41%), Positives = 169/281 (59%), Gaps = 5/281 (1%)

Query: 1   MNKK--KRLNLTFVYILLIVLSIMWLFPIVWVVLTSFRGEGSAFVNYFIPKTWTLDNYAK 58
           MNKK   RL +T +Y+ L+V+ I+ L+P++W V  S     S F +  IP+T +  +Y
Sbjct: 1   MNKKVKSRLEVTAIYLFLLVMGIVILYPLLWTVGLSLNPGTSLFSSRMIPETISFRHYEW 60

Query: 59  LFTQNTFPFGQWFLNTLFVATCTCILSTLITVAMAYSLSRIKFKHRNGFLKLALVLNMFP 118
           LF      + QW+ NTL VA+ T + ST +    AY+ SR +F  R   L   L+L MFP
Sbjct: 61  LFFDPRSNYLQWYKNTLIVASVTSVCSTFLVALTAYAFSRYRFVGRTYGLYGFLLLQMFP 120

Query: 119 GFMSMIAVYYILKALNLDQTLTALIFVY-SAGAALTFYIAKGFFDTIPYSLDESAMIDGA 177
             M+M+A+Y +L  +NL  TL  LI +Y      +  ++ KG+FDTIP  LDESA +DGA
Sbjct: 121 VLMAMVALYILLNTVNLLDTLLGLILIYVGTSIPMNAFLVKGYFDTIPRELDESAKLDGA 180

Query: 178 TRLDIFLKITLPLSKPIIVYTALIAFMGPWMDFIFAKVILGDATSKYTVAIGLFSMLQQD 237
               IF  I LPL+KPI+   AL  FM P+MDFI  ++IL  +   YT+A+GLF+ +
Sbjct: 181 GHFRIFFTIMLPLAKPILAVVALFNFMSPFMDFILPRIIL-RSPENYTLALGLFNFVNDQ 239

Query: 238 TINQWFMSFTAGSVIIAIPITILFMFMQKYYVEGITGGSVK 278
             N  F  F AG+++IAIPI  +F+F+Q+Y + G+T G+ K
Sbjct: 240 FANN-FTRFAAGAILIAIPIATVFLFLQRYLISGLTTGATK 279
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5069> which encodes the amino acid sequence <SEQ ID 5070>. Analysis of this protein sequence reveals the following:

```
        Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood = -6.42   Transmembrane   76 -  92 (  71 -  97)
     INTEGRAL    Likelihood = -6.05   Transmembrane  248 - 264 ( 242 - 267)
     INTEGRAL    Likelihood = -3.50   Transmembrane  110 - 126 ( 110 - 127)
     INTEGRAL    Likelihood = -1.33   Transmembrane  129 - 145 ( 129 - 145)
     INTEGRAL    Likelihood = -1.33   Transmembrane  188 - 204 ( 188 - 204)

----- Final Results -----
                 bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA60006 GB:X86014 cymG [Klebsiella oxytoca]
 Identities = 119/270 (44%), Positives = 172/270 (63%), Gaps = 7/270 (2%)

Query: 11  LVYATLIILSIIWLFPIAWVILTSFRSEGTAYVNYFIPKTFTLNHYINLFTNETFPFGKW 70
           LVY  L++ +++ L P+ W +++S +     + + F    +FTL HY NL T    P+ KW
Sbjct: 12  LVYLFLLLNALVVLGPVIWTVMSSLKPGNNLFSSGFTEISFTLEHYHNLLTGT--PYLKW 69

Query: 71  FMNTLIVATFTCIISTFITVAIAYSLSRIKFKFRNGFLKLALILNMFPGFMSMIAIYYIL 130
           + NT I+AT   +IS  +    A+  SR +FK +   L   L+L MFP F+SM AIY +L
Sbjct: 70  YKNTFILATCNMLISLVVVTITAFIFSRYRFKAKKKILMSILVLQMFPAFLSMTAIYILL 129

Query: 131 KALGLTQTLTALVLVYSSGAALGF--YIAKGFFDTIPYSLDESAMIDGATRMDIFFKITL 188
             + L  T   L+LVY +G+ L F  ++ KG+FD IP SLDE+A IDGA  + IFF+I L
Sbjct: 130 SKMNLIDTYIGLLLVYVTGS-LPFMTWLVKGYFDAIPTSLDEAAKIDGAGHLTIFFEIIL 188

Query: 189 PLAKPIIVYTALLAFMGPWIDFIFAQVILGDATSKYTVAIGLFSMLQPDTINNWFMAFTA 248
```

```
           PLAKPI+V+ AL++F GPW+DFI   +IL  +  K T+AIG+FS +  ++  N F  F A
Sbjct: 189 PLAKPILVFVALVSFTGPWMDFILPTLIL-RSEDKMTLAIGIFSWISSNSAEN-FTLFAA 246

Query: 249 GSVLIAVPITLLFMFMQKYYVEGITGGSVK 278
           G++L+AVPITLLF+  QK+   G+  G+VK
Sbjct: 247 GALLVAVPITLLFIVTQKHITTGLVSGAVK 276
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 227/278 (81%), Positives = 253/278 (90%)

Query: 1   MNKKKRLNLTFVYILLIVLSIMWLFPIVWVVLTSFRGEGSAFVNYFIPKTWTLDNYAKLF 60
           M  K+R  L  VY  LI+LSI+WLFPI WV+LTSFR EG+A+VNYFIPKT+TL++Y  LF
Sbjct: 1   MKNKRRFQLGLVYATLIILSIIWLFPIAWVILTSFRSEGTAYVNYFIPKTFTLNHYINLF 60

Query: 61  TQNTFPFGQWFLNTLFVATCTCILSTLITVAMAYSLSRIKFKHRNGFLKLALVLNMFPGF 120
           T  TFPFG+WF+NTL VAT TCI+ST ITVA+AYSLSRIKFK RNGFLKLAL+LNMFPGF
Sbjct: 61  TNETFPFGKWFMNTLIVATFTCIISTFITVAIAYSLSRIKFKFRNGFLKLALILNMFPGF 120

Query: 121 MSMIAVYYILKALNLDQTLTALIFVYSAGAALTFYIAKGFFDTIPYSLDESAMIDGATRL 180
           MSMIA+YYILKAL L QTLTAL+ VYS+GAAL FYIAKGFFDTIPYSLDESAMIDGATR+
Sbjct: 121 MSMIAIYYILKALGLTQTLTALVLVYSSGAALGFYIAKGFFDTIPYSLDESAMIDGATRM 180

Query: 181 DIFLKITLPLSKPIIVYTALIAFMGPWMDFIFAKVILGDATSKYTVAIGLFSMLQQDTIN 240
           DIF KITLPL+KPIIVYTAL+AFMGPW+DFIFA+VILGDATSKYTVAIGLFSMLQ DTIN
Sbjct: 181 DIFFKITLPLAKPIIVYTALLAFMGPWIDFIFAQVILGDATSKYTVAIGLFSMLQPDTIN 240

Query: 241 QWFMSFTAGSVIIAIPITILFMFMQKYYVEGITGGSVK 278
            WFM+FTAGSV+IA+PIT+LFMFMQKYYVEGITGGSVK
Sbjct: 241 NWFMAFTAGSVLIAVPITLLFMFMQKYYVEGITGGSVK 278
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1640

A DNA sequence (GBSx1735) was identified in *S.agalactiae* <SEQ ID 5071> which encodes the amino acid sequence <SEQ ID 5072>. This protein is predicted to be cymF protein (malF). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.46   Transmembrane  427 - 443 ( 417 - 447)
   INTEGRAL    Likelihood =-10.24   Transmembrane   99 - 115 (  96 - 121)
   INTEGRAL    Likelihood = -9.39   Transmembrane  166 - 182 ( 154 - 185)
   INTEGRAL    Likelihood = -6.21   Transmembrane  259 - 275 ( 257 - 276)
   INTEGRAL    Likelihood = -6.21   Transmembrane  229 - 245 ( 223 - 247)
   INTEGRAL    Likelihood = -6.10   Transmembrane   44 -  60 (  40 -  66)
   INTEGRAL    Likelihood = -4.51   Transmembrane  314 - 330 ( 312 - 331)

----- Final Results -----
               bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60005 GB:X86014 cymF [Klebsiella oxytoca]
 Identities = 174/428 (40%), Positives = 263/428 (60%), Gaps = 21/428 (4%)

Query: 27   SFLIMGLANLKNKQIVKGLLFLISEILFLITFVYQVIPAVKGLISLGTQEQGMTTKTVDG 86
            SFLIMG   L +   +KG +FL+ +I+ +I+ +  ++ A +GLI+LGT  Q     T  G
Sbjct: 15   SFLIMGATQLISGHWIKGSVFLLFQIV-VISNINLLLNATQGLITLGTVAQ-----TRSG 68

Query: 87   IKIQVATQGDNSMLMLIFGLASLIFCCVFAYIYWSNIKSAAHLLTLKEEGREIPSFKKDI 146
```

```
             I     GDNS+ ML+ G+ + IF     ++YW NIK A       +      SF + +
Sbjct: 69  FDI---VAGDNSIFMLVEGVVAFIFLFFSIFVYWLNIKDAQVCEKCHQ------SFTEQL 119

Query: 147 KSLTDGRFHMTLMSIPLIGVLLFTILPLVYMICLAFTNYDH-NHLPPKSLFDWVGFANFG 205
           +++ D RF  +++    I  + F I+P++  + ++ TNY   +H+PPK+L DWVG  NF
Sbjct: 120 RTIYDNRFATIMLAPAFIACIAFIIMPMIITVLVSLTNYSAPHHIPPKNLVDWVGLKNFI 179

Query: 206 NIFSGRMAS-TFFPVLSWTLIWAVFATVTNFFFGIILALLINTKGLKFKKMWRTIFVITM 264
            +F  R+ S TF  +  WT++WA FAT+    FG +LAL +  K +  KK WR +F++
Sbjct: 180 TLFELRIWSKTFVGIGVWTVLWAFFATLCTCSFGFLLALALENKKIIAKKAWRVVFILPY 239

Query: 265 AVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHPLPFLSDPVWAKFSIIFVNMWVGIP 324
           A+P F++LLI R LL+  GPVN+ L  WG+ S    + FLSDP+ AK ++I V++WVG P
Sbjct: 240 AIPAFVTLLIFRLLLNGIGPVNSTLNSWGIDS----IGFLSDPLIAKMTVIAVSVWVGAP 295

Query: 325 VTMLVATGIIMNLPAEQIEAAEIDGANKFQVFQSITFPQILLIMTPTLIQQFIGNINNFN 384
             ML+ TG + N+P +  EA+E+DGA+KFQ F+ IT P +L  + P+L+  F  N NNF
Sbjct: 296 YFMLLITGAMTNIPRDLYEASEVDGASKFQQFREITLPMVLHQVAPSLVMTFAHNFNNFG 355

Query: 385 VIYLLTQGGPTNSTYYQAGSTDLLVTWLYNLTVTAADYNLASVVGILIFILSAVFSLLAY 444
            IYLLT+GGP N  Y  AG TD+L+TW+Y LT+    Y +ASV+ I+IF+  ++F++  +
Sbjct: 356 AIYLLTEGGPINPEYRFAGHTDILITWIYKLTLDFQQYQIASVISIIIFLFLSIFAIWQF 415

Query: 445 TRTNSYKE 452
            R  S+KE
Sbjct: 416 RRMKSFKE 423
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5073> which encodes the amino acid sequence <SEQ ID 5074>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.93   Transmembrane   98 - 114 (  95 - 122)
   INTEGRAL    Likelihood = -9.55   Transmembrane  165 - 181 ( 152 - 184)
   INTEGRAL    Likelihood = -9.24   Transmembrane  424 - 440 ( 419 - 443)
   INTEGRAL    Likelihood = -7.91   Transmembrane   43 -  59 (  39 -  71)
   INTEGRAL    Likelihood = -7.59   Transmembrane  258 - 274 ( 256 - 275)
   INTEGRAL    Likelihood = -6.21   Transmembrane  228 - 244 ( 222 - 246)
   INTEGRAL    Likelihood = -4.09   Transmembrane  311 - 327 ( 309 - 328)

----- Final Results -----
               bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA60005 GB:X86014 cymF [Klebsiella oxytoca]
 Identities = 179/426 (42%), Positives = 266/426 (62%), Gaps = 19/426 (4%)

Query: 26  SSIIMGFANFANKQFIKGILFLISELIFLVAFVSQIIPAIRGLVTLGTQTQGMTTKTIDG 85
           S +IMG     +  +IKG +FL+ +++ +++ ++ ++ A +GL+TLGT  Q     T  G
Sbjct: 15  SFLIMGATQLISGHWIKGSVFLLFQIV-VISNINLLLNATQGLITLGTVAQ-----TRSG 68

Query: 86  INIQVAVDGDNSMLMLIFGLASLIFCLVFAYIYWCNLKSARNLYLFKQKGQKIPSFKEDL 145
            +I   V GDNS+ ML+ G+ + IF     ++YW N+K A+      Q      SF E L
Sbjct: 69  FDI---VAGDNSIFMLVEGVVAFIFLFFSIFVYWLNIKDAQVCEKCHQ------SFTEQL 119

Query: 146 ATLTNGRFHMTLMAIPLIGVLLFTILPLIYMICLAFTNFDH-NHLPPKSLFDWVGLANFG 204
            T+ + RF   ++A   I  + F I+P+I  + ++ TN+   +H+PPK+L DWVGL NF
Sbjct: 120 RTIYDNRFATIMLAPAFIACIAFIIMPMIITVLVSLTNYSAPHHIPPKNLVDWVGLKNFI 179

Query: 205 NVLSGRM-AGTFFPIFSWTLIWAVFATVTNFFFGIILALLINTKGLKWKKMWRTIFVITI 263
            +   R+ + TF  I  WT++WA FAT+    FG +LAL +  K +  KK WR +F++
Sbjct: 180 TLFELRIWSKTFVGIGVWTVLWAFFATLCTCSFGFLLALALENKKIIAKKAWRVVFILPY 239

Query: 264 AVPQFISLLIMRNLLNDEGPLNALLNKIGLINGSLPFLSDPLWAKFSIIFVNMWIGIPFT 323
```

```
           A+P F++LLI R LLN  GP+N+ LN  G+   S+ FLSDPL AK ++I V++W+G P+
Sbjct: 240 AIPAFVTLLIFRLLLNGIGPVNSTLNSWGI--DSIGFLSDPLIAKMTVIAVSVWVGAPYF 297

Query: 324 MLIATGIIMNLPSEQIEAAEIDGASKFQVFKSITFPQILLIMTPNLIQQFIGNINNFNVI 383
           ML+ TG + N+P +  EA+E+DGASKFQ F+ IT P +L  + P+L+  F  N NNF  I
Sbjct: 298 MLLITGAMTNIPRDLYEASEVDGASKFQQFREITLPMVLHQVAPSLVMTFAHNFNNFGAI 357

Query: 384 YLLTGGGPTNSEYYQAGTTDLLVTWLYKLTVTAADYNLASVIGILIFTVSAIFSLLAYTR 443
           YLLT GGP N EY  AG TD+L+TW+YKLT+    Y +ASVI I+IF   +IF++  + R
Sbjct: 358 YLLTEGGPINPEYRFAGHTDILITWIYKLTLDFQQYQIASVISIIIFLFLSIFAIWQFRR 417

Query: 444 TASYKE 449
             S+KE
Sbjct: 418 MKSFKE 423
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 357/446 (80%), Positives = 404/446 (90%), Gaps = 2/446 (0%)

Query: 11  MSLKEVFQKGDLATKLSFLIMGLANLKNKQIVKGLLFLISEILFLITFVYQVIPAVKGLI 70
           +S+ E  ++G    KLS +IMG AN  NKQ +KG+LFLISE++FL+ FV Q+IPA++GL+
Sbjct: 10  ISVIEALKRGSWDIKLSSIIMGFANFANKQFIKGILFLISELIFLVAFVSQIIPAIRGLV 69

Query: 71  SLGTQEQGMTTKTVDGIKIQVATQGDNSMLMLIFGLASLIFCCVFAYIYWSNIKSAAHLL 130
           +LGTQ QGMTTKT+DGI IQVA  GDNSMLMLIFGLASLIFC VFAYIYW N+KSA +L
Sbjct: 70  TLGTQTQGMTTKTIDGINIQVAVDGDNSMLMLIFGLASLIFCLVFAYIYWCNLKSARNLY 129

Query: 131 TLKEEGREIPSFKKDIKSLTDGRFHMTLMSIPLIGVLLFTILPLVYMICLAFTNYDHNHL 190
             K++G++IPSFK+D+ +LT+GRFHMTLM+IPLIGVLLFTILPL+YMICLAFTN+DHNHL
Sbjct: 130 LFKQKGQKIPSFKEDLATLTNGRFHMTLMAIPLIGVLLFTILPLIYMICLAFTNFDHNHL 189

Query: 191 PPKSLFDWVGFANFGNIFSGRMASTFFPVLSWTLIWAVFATVTNFFFGIILALLINTKGL 250
           PPKSLFDWVG ANFGN+ SGRMA TFFP+ SWTLIWAVFATVTNFFFGIILALLINTKGL
Sbjct: 190 PPKSLFDWVGLANFGNVLSGRMAGTFFPIFSWTLIWAVFATVTNFFFGIILALLINTKGL 249

Query: 251 KFKKMWRTIFVITMAVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHPLPFLSDPVWA 310
           K+KKMWRTIFVIT+AVPQFISLLIMRNLL+D GP+NALL K GLI+ +  LPFLSDP+WA
Sbjct: 250 KWKKMWRTIFVITIAVPQFISLLIMRNLLNDEGPLNALLNKIGLINGS--LPFLSDPLWA 307

Query: 311 KFSIIFVNMWVGIPVTMLVATGIIMNLPAEQIEAAEIDGANKFQVFQSITFPQILLIMTP 370
           KFSIIFVNMW+GIP TML+ATGIIMNLP+EQIEAAEIDGA+KFQVF+SITFPQILLIMTP
Sbjct: 308 KFSIIFVNMWIGIPFTMLIATGIIMNLPSEQIEAAEIDGASKFQVFKSITFPQILLIMTP 367

Query: 371 TLIQQFIGNINNFNVIYLLTQGGPTNSTYYQAGSTDLLVTWLYNLTVTAADYNLASVVGI 430
            LIQQFIGNINNFNVIYLLT GGPTNS YYQAG+TDLLVTWLY LTVTAADYNLASV+GI
Sbjct: 368 NLIQQFIGNINNFNVIYLLTGGGPTNSEYYQAGTTDLLVTWLYKLTVTAADYNLASVIGI 427

Query: 431 LIFILSAVFSLLAYTRTNSYKEGAAK 456
           LIF +SA+FSLLAYTRT SYKEGAAK
Sbjct: 428 LIFTVSAIFSLLAYTRTASYKEGAAK 453
```

A related GBS gene <SEQ ID 8869> and protein <SEQ ID 8870> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:    -12.73
GvH: Signal Score (-7.5): -6.04
     Possible site: 36
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 7 value: -11.46 threshold:  0.0
   INTEGRAL    Likelihood =-11.46   Transmembrane  427 - 443 ( 417 - 447)
   INTEGRAL    Likelihood = -9.87   Transmembrane   99 - 115 (  96 - 121)
   INTEGRAL    Likelihood = -9.39   Transmembrane  166 - 182 ( 154 - 185)
   INTEGRAL    Likelihood = -6.21   Transmembrane  259 - 275 ( 257 - 276)
   INTEGRAL    Likelihood = -6.21   Transmembrane  229 - 245 ( 223 - 247)
   INTEGRAL    Likelihood = -6.10   Transmembrane   44 -  60 (  40 -  66)
   INTEGRAL    Likelihood = -4.51   Transmembrane  314 - 330 ( 312 - 331)
```

```
   PERIPHERAL  Likelihood =  0.90     212
 modified ALOM score:   2.79

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01027(379 - 1656 of 1968)
EGAD|33392|34706(15 - 423 of 427) cymF protein {Klebsiella oxytoca}
GP|854233|emb|CAA60005.1||X86014 cymF {Klebsiella oxytoca} PIR|S63615|S63615 malF protein
homolog cymF - Klebsiella oxytoca
%Match = 23.8
%Identity = 41.3  %Similarity = 64.5
Matches = 171  Mismatches = 140  Conservative Sub.s = 96

132       162       192       222       252       282       312       342
VLLFLAILTVVKSNLAITLNV*NNSIKTSLKQNSTSRVMR*GEYSSFQLRVLPISYFVK*QLKETIMNKKLISLDGMSLK

                                                                              ML

372       402       432       462       492       522       552       582
EVFQKGDLATKLSFLIMGLANLKNKQIVKGLLFLISEILFLITFVYQVIPAVKGLISLGTQEQGMTTKTVDGIKIQVATQ
            ||||||   | : : :|| :||: :|  :|: :  :: | :|||:|||  |     |  |  |
LSEGKSMRIFPASFLIMGATQLISGHWIKGSVFLLFQI-VVISNINLLLNATQGLITLGTVAQ-----TRSGFDI---VA
                20        30        40         50        60              70

612       642       672       702       732       762       792       822
GDNSMLMLIFGLASLIFXCVFAYIYWSNIXSAAHLLTLKEEGREIPSFKKDIKSLTDGRFHMTLMSIPLIGVLLFTILPL
||||::||: |: ::||     ::|| ||         :   :   || : :::: | ||   :::  :|  : | |:|:
GDNSIFMLVEGVVAFIFLFFSIFVYWLNI------KDAQVCEKCHQSFTEQLRTIYDNRFATIMLAPAFIACIAFIIMPM
                90       100             110       120       130       140

852       879       909       939       966       996       1026      1056
VYMICLAFTNYDH-NHLPPKSLFDWVGFANFGNIFSGRMAS-TFFPVLSWTLIWAVFATVTNFFFGIILALLINTKGLKF
:  : :::|||   :|:|||:| ||||: ||  :|  |: | ||  :  ||::|| |||:    || :||| :  | :
IITVLVSLTNYSAPHHIPPKNLVDWVGLKNFITLFELRIWSKTFVGIGVWTVLWAFFATLCTCSFGFLLALALENKKIIA
           160       170       180       190       200       210       220

1086      1116      1146      1176      1206      1236      1266      1296
KKMWRTIFVITMAVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHPLPFLSDPVWAKFSIIFVNMWVGIPVTMLVATG
|| || :|::  |:| |::||| | ||:  ||||: |  ||       : |||||: || ::| |::||| |  ||: ||
KKAWRVVFILPYAIPAFVTLLIFRLLLNGIGPVNSTLNSWG----IDSIGFLSDPLIAKMTVIAVSVWVGAPYFMLLITG
           240       250       260           270       280       290       300

1326      1356      1386      1416      1446      1476      1506      1536
IIMNLPAEQIEAAEIDGANKFQVFQSITFPQILLIMTPTLIQQFIGNINNFNVIYLLTQGGPTNSTYYQAGSTDLLVTWL
 : |:| :  ||:|:|||:||| |: ||:| :|  : |:|:  |  | |||  |||||:||| |  |  || ||:|:||:
AMTNIPRDLYEASEVDGASKFQQFREITLPMVLHQVAPSLVMTFAHNFNNFGAIYLLTEGGPINPEYRFAGHTDILITWI
               320       330       340       350       360       370       380

1566      1596      1626      1656      1686      1716      1746      1776
YNLTVTAADYNLASVVGILIFILSAVFSLLAYTRTNSYKEGAAK**IRKNVLTLLLFIFY**YYQLCGSFPLFGSFSQAS
| ||:    | :|||: |:||:: ::|::  : |  |:||
YKLTLDFQQYQIASVISIIIFLFLSIFAIWQFRRMKSFKEDVGM
               400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1641

A DNA sequence (GBSx1736) was identified in *S.agalactiae* <SEQ ID 5075> which encodes the amino acid sequence <SEQ ID 5076>. This protein is predicted to be maltose/maltodextrin-binding protein precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.98    Transmembrane   25 -  41 (  24 -  43)

----- Final Results -----
              bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9999> which encodes amino acid sequence <SEQ ID 10000> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA26925 GB:L08611 MalX [Streptococcus pneumoniae]
 Identities = 117/418 (27%), Positives = 186/418 (43%), Gaps = 43/418 (10%)

Query: 15  TKMEKNTWKKLLVSTAALSVVAGGAIAATHSNSVDAASKTTIKLWVPTDSKASYKAIVKK 74
            +K  K+T     V+ A+L +VA G+  A       ++    + ++V    K+  + + K
Sbjct: 3   SKFMKSTAVLGTVTLASLLLVACGSKTADKPADSGSSEVKELTVYVDEGYKSYIEEVAKA 62

Query: 75  FZKE-NKGVTVKMIESNDSKAQENVKKDPSKAADVFSLPHDQLGQLVESGVIQEIPEQYS 133
            ++KE    VT+K  ++     + ++        DV   P+D++G L   G + E+  + S
Sbjct: 63  YEKEAGVKVTLKTGDALGGLDKLSLDNQSGNVPDVMMAPYDRVGSLGSDGQLSEV--KLS 120

Query: 134 KEIAKNDTKQSLTGAQYKGKTYAFPFGIESQVLYYNKTKLTADDVKSYETITSKGKFGXQ 193
                 +DT +SL  A   GK Y  P  IES V+YYNK  L  D  K++  + +  K
Sbjct: 121 DGAKTDDTTKSLVTAA-NGKVYGAPAVIESLVMYYNKD-LVKDAPKTFADLENLAKDSKY 178

Query: 194 LKAA-------------NSYVTGPXFLSVGDTLFGKSGEDAKGTNWGNEAGVSVL----- 235
              A              N Y T       G  +FG++G+DAK     N+  ++ +
Sbjct: 179 AFAGEDGKTTAFLADWTNFYYTYGLLAGNGAYVFGQNGKDAKDIGLANDGSIAGINYAKS 238

Query: 236 ---KWIADQKKNDGFVNLTAENTMSKFGDGSVHAFESGPWDYDAAKKAVGEDKIGVAVYP 292
               KW    +  +G  NL      ++F +G   A   GPW   A K A  +   GVA  P
Sbjct: 239 WYEKWPKGMQDTEGAGNLI----QTQFQEGKTAAIIDGPWKAQAFKDA--KVNYGVATIP 292

Query: 293 TMKIGDKEVQQKAFLGVKLYAVNQAPAGSNTKRISASYKLAAYLTNAESQKIQFEKRHIV 352
            T+  G    +  AF G K + + QA      K + AS  K    +L   E QK+  ++K + +
Sbjct: 293 TLPNGK---EYAAFGGGKAWVIPQA-----VKNLEASQKFVDFLVATEQQKVLYDKTNEI 344

Query: 353 PANSSIQSSDSVQKDELAKAVIEMGSSDKYTTVMPKLSQMSTFWTESAAILSDTYSGK 410
            PAN+  +S    + DEL  AVI+     K T  +P +SQMS  W  +  +L D  SG+
Sbjct: 345 PANTEARSYAEGKNDELTTAVIK---QFKNTQPLPNISQMSAVWDPAKNMLFDAVSGQ 399
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5077> which encodes the amino acid sequence <SEQ ID 5078>. Analysis of this protein sequence reveals the following:

```
     Possible site: 28

>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA26925 GB:L08611 MalX [Streptococcus pneumoniae]
 Identities = 126/423 (29%), Positives = 191/423 (44%), Gaps = 50/423 (11%)

Query: 13  SLTLASTLLVGCGSGSKDK--KEAGADSKTIKLWVPTGSKKSYADTIAK-FEKDSGYTVK 69
            ++TLAS LLV CGS + DK       ++K + ++V  G  KSY + +AK +EK++G  V
Sbjct: 14  TVTLASLLLVACGSKTADKPADSGSSEVKELTVYVDEG-YKSYIEEVAKAYEKEAGVKVT 72

Query: 70  VVESEDPKAQEKIKKD--ASTAADVFSLPHDQLGQLVESGTIQEVPEKYNKEIAATSTDQ 127
            +   +     +K+  D  +    DV   P+D++G L   G + EV  K +       T +
Sbjct: 73  LKTGDALGGLDKLSLDNQSGNVPDVMMAPYDRVGSLGSDGQLSEV--KLSDGAKTDDTTK 130

Query: 128 ALVGAQYKGKTYAFPFGIESQVLFYNKSKLAAEDVTSYD----TITTKATFGGTFKQ--- 180
            +LV A   GK Y  P  IES V++YNK  +     T  D       +K  F G   +
Sbjct: 131 SLVTAA-NGKVYGAPAVIESLVMYYNKDLVKDAPKTFADLENLAKDSKYAFAGEDGKTTA 189

Query: 181 -----ANTYATGPLFMSVGNTLFGENGEDVKGTNWGNEKGAAVL--------KWIADQAS 227
                  N Y T  L    G  +FG+NG+D K     N+   A +        KW
Sbjct: 190 FLADWTNFYYTYGLLAGNGAYVFGQNGKDAKDIGLANDGSIAGINYAKSWYEKWPKGMQD 249

Query: 228 NKGFVSLDANNVMSKFGDGSVASFESGPWDYEAAQKAIGKENLGVAIYPKVTIGGETVQQ 287
             +G      N + ++F +G  A+   GPW  +A + A  K N GVA  P +  G E
Sbjct: 250 TEG----AGNLIQTQFQEGKTAAIIDGPWKAQAFKDA--KVNYGVATIPTLPNGKE---Y 300

Query: 288 KAFLGVKLYAVNQAPAKGDTKRIAASYKLASYLTNAESQENQFKTRNIVPANKEVQSSEA 347
             AF G K + + QA      K + AS K   +L   E Q+  +   N +PAN E +S
Sbjct: 301 AAFGGGKAWVIPQA-----VKNLEASQKFVDFLVATEQQKVLYDKTNEIPANTEARSYAE 355

Query: 348 VQSNELAKTVITMGSSSDYTVVMPKLSQMGTFWTESAAILSDAFNG----KIKENDYLTK 403
             +++EL   VI    +   T  +P +SQM   W  +  +L DA +G    K   ND +T
Sbjct: 356 GKNDELTTAVIKQFKN---TQPLPNISQMSAVWDPAKNMLFDAVSGQKDAKTAANDAVTL 412

Query: 404 LQQ 406
            +++
Sbjct: 413 IKE 415
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 278/415 (66%), Positives = 334/415 (79%), Gaps = 6/415 (1%)

Query: 21  TWKKLLVSTAALSVVAGGAIAATHSNSVD----AASKTTIKLWVPTDSKASYKAIVKKFZ 76
            +W+K++V  A+L++ A   +    S S D     A   TIKLWVPT SK SY   + KF+
Sbjct: 3   SWQKVIVGGASLTL-ASTLLVGCGSGSKDKKEAGADSKTIKLWVPTGSKKSYADTIAKFE 61

Query: 77  KENKGVTVKMIESNDSKAQENVKKDPSKAADVFSLPHDQLGQLVESGVIQEIPEQYSKEI 136
            K++ G TVK++ES D KAQE +KKD S AADVFSLPHDQLGQLVESG IQE+PE+Y+KEI
Sbjct: 62  KDS-GYTVKVVESEDPKAQEKIKKDASTAADVFSLPHDQLGQLVESGTIQEVPEKYNKEI 120

Query: 137 AKNDTKQSLTGAQYKGKTYAFPFGIESQVLYYNKTKLTADDVKSYETITSKGKFGXQLKA 196
            A   T Q+L GAQYKGKTYAFPFGIESQVL+YNK+KL A+DV SY+TIT+K  FG   K
Sbjct: 121 AATSTDQALVGAQYKGKTYAFPFGIESQVLFYNKSKLAAEDVTSYDTITTKATFGGTFKQ 180

Query: 197 ANSYVTGPXFLSVGDTLFGKSGEDAKGTNWGNEAGVSVLKWIADQKKNDGFVNLTAENTM 256
            AN+Y TGP F+SVG+TLFG++GED KGTNWGNE G +VLKWIADQ  N GFV+L A N M
Sbjct: 181 ANTYATGPLFMSVGNTLFGENGEDVKGTNWGNEKGAAVLKWIADQASNKGFVSLDANNVM 240

Query: 257 SKFGDGSVHAFESGPWDYDAAKKAVGEDKIGVAVYPTMKIGDKEVQQKAFLGVKLYAVNQ 316
            SKFGDGSV +FESGPWDY+AA+KA+G++ +GVA+YP + IG + VQQKAFLGVKLYAVNQ
Sbjct: 241 SKFGDGSVASFESGPWDYEAAQKAIGKENLGVAIYPKVTIGGETVQQKAFLGVKLYAVNQ 300

Query: 317 APAGSNTKRISASYKLAAYLTNAESQKIQFEKRHIVPANSSIQSSDSVQKDELAKAVIEM 376
            APA  +TKRI+ASYKLA+YLTNAESQ+ QF+ R+IVPAN  +QSS++VQ +ELAK VI M
Sbjct: 301 APAKGDTKRIAASYKLASYLTNAESQENQFKTRNIVPANKEVQSSEAVQSNELAKTVITM 360

Query: 377 GSSDKYTTVMPKLSQMSTFWTESAAILSDTYSGKIKSSDYLKRLKQFDKDIAKTK 431
            GSS  YT VMPKLSQM TFWTESAAILSD ++GKIK +DYL +L+QFDKDIA TK
Sbjct: 361 GSSSDYTVVMPKLSQMGTFWTESAAILSDAFNGKIKENDYLTKLQQFDKDIAATK 415
```

SEQ ID 5076 (GBS649) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 2 & 3; MW 76kDa) and in Figure 186 (lane 7; MW 76kDa).. It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 7; MW 51kDa) and in Figure 178 (lane 8; MW 51kDa).

GBS649-His was purified as shown in Figure 229, lane 8. Purified GBS649-GST is shown in Figure 245, lanes 6 &73.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1642

A DNA sequence (GBSx1737) was identified in *S.agalactiae* <SEQ ID 5079> which encodes the amino acid sequence <SEQ ID 5080>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2462(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD02112 GB:AF039082 putative maltose operon transcriptional
          repressor [Lactococcus lactis]
 Identities = 43/61 (70%), Positives = 49/61 (79%)

Query: 2  VTIKDVAAKAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTQN 61
           VTIKDVA KAGVN STVSRV+KD++ IS KTK +V+KAM ELGY  N AAQ+LASG T
Sbjct: 3  VTIKDVAKKAGVNASTVSRVIKDSSEISDKTKVKVRKAMHELGYRRNAAAQILASGKTNT 62

Query: 62 I 62
          I
Sbjct: 63 I 63
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5081> which encodes the amino acid sequence <SEQ ID 5082>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.93   Transmembrane  269 - 285 ( 266 - 287)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 53/62 (85%), Positives = 57/62 (91%)

Query: 1  MVTIKDVAAKAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTQ 60
           MVTIKDVA KAGVNPSTVSRVLKDN SIS KTKE+V+KAM +LGYVPNVAAQ+LASGLT
Sbjct: 26 MVTIKDVAQKAGVNPSTVSRVLKDNRSISMKTKEKVRKAMADLGYVPNVAAQILASGLTH 85

Query: 61 NI 62
           NI
Sbjct: 86 NI 87
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1643

A DNA sequence (GBSx1738) was identified in *S.agalactiae* <SEQ ID 5083> which encodes the amino acid sequence <SEQ ID 5084>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.70   Transmembrane   14 -  30 (   8 -  34)
   INTEGRAL    Likelihood = -6.90   Transmembrane   66 -  82 (  63 -  85)
   INTEGRAL    Likelihood = -6.69   Transmembrane  110 - 126 ( 105 - 128)
   INTEGRAL    Likelihood = -3.93   Transmembrane  132 - 148 ( 129 - 149)

----- Final Results -----
               bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9443> which encodes amino acid sequence <SEQ ID 9444> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67260 GB:AF017113 YvjA [Bacillus subtilis]
 Identities = 83/227 (36%), Positives = 140/227 (61%)

Query: 9   FGWDSAFFIMIINIPLLLLCYFGLGKQTFLKTVYGSWIFPVFIKLTQSVPTLTHNPLLAA 68
           +G+++A+   IINIPL +     LG +  LKT+ GS   P+ + LT+ +   TH+ LLAA
Sbjct: 52  YGFEAAYVQWIINIPLFIAGVILLGGKFGLKTLAGSVFLPLVVFLTRDIQPATHHELLAA 111

Query: 69  LFGGVIVGCGLGIVFWSDSSTGGTGIIIQFLGKYTPISLGQGVILIDGLVTIVGFLAFDS 128
           +FGGV +G G+GIV+    STGGT +  Q + KY+ +SLG+ + +IDG++ +    + F+
Sbjct: 112 IFGGVGIGIGIGIVYLGKGSTGGTALAAQIIHKYSGLSLGKCLAIIDGMIVVTAMIVFNI 171

Query: 129 DTVMFSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGVTEIPVKGG 188
           +  +++++G+   S  I+ +Q GF      LI++++ Q +K  +    DRGVT+I   GG
Sbjct: 172 EQGLYAMLGVYVSSKTIDVVQVGFNRSKMALIITKQEQAVKEAVLQKIDRGVTKISAVGG 231

Query: 189 YSGTNQIMLMTTIAGYEFAKLQEAIAEIDETAFITVTPTSQASGRGF 235
           Y+  ++ +LM  +   EF KL++ + +IDE+AF+ V   S+  G GF
Sbjct: 232 YTDDDRPILMCVVGQTEFTKLKQIVKQIDESAFVIVADASEVLGEGF 278
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5085> which encodes the amino acid sequence <SEQ ID 5086>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.21   Transmembrane  104 - 120 ( 101 - 123)
   INTEGRAL    Likelihood = -3.93   Transmembrane  147 - 163 ( 142 - 167)
   INTEGRAL    Likelihood = -3.29   Transmembrane  169 - 185 ( 169 - 186)

----- Final Results -----
               bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC67260 GB:AF017113 YvjA [Bacillus subtilis]
 Identities = 106/267 (39%), Positives = 169/267 (62%), Gaps = 1/267 (0%)

Query: 7   DLLLVTIGSFITAIGFNTMFVDNHIASGGMVGIAVVIKALFGISPSLFLMASNIPLLLMC 66
```

```
           D + + IG+ ITA+ FN   + N IA+GG+ GI+ ++++ +G   +      NIPL +
Sbjct: 13  DYVYILIGAAITAVSFNVFLLPNKIAAGGVSGISTILQS-YGFEAAYVQWIINIPLFIAG 71

Query: 67  YFFLGKQNFIKTLYGSWIYPIAIRSTNSLPTLTHNQLLAAIFGGIICGIGLGMVFWGNSS 126
              LG +  +KTL GS   P+ +  T  +   TH++LLAAIFGG+  GIG+G+V+ G  S
Sbjct: 72  VILLGGKFGLKTLAGSVFLPLVVFLTRDIQPATHHELLAAIFGGVGIGIGIGIVYLGKGS 131

Query: 127 TGGTGILTQILHKYSPLSLGVAMTIVDGISVLMGFIALSADDVMYSTIGLFVIGYVISVM 186
           TGGT +  QI+HKYS LSLG  + I+DG+ V+   I  + +  +Y+ +G++V    I V+
Sbjct: 132 TGGTALAAQIIHKYSGLSLGKCLAIIDGMIVVTAMIVFNIEQGLYAMLGVYVSSKTIDVV 191

Query: 187 ENGFDSSKNVMIISKDYQAIREYITTVMDRGVTKLPIRGGYTTSDKIMLMAIVSSHELPT 246
           + GF+ SK  +II+K  QA++E +   +DRGVTK+   GGYT  D+ +LM +V   E
Sbjct: 192 QVGFNRSKMALIITKQEQAVKEAVLQKIDRGVTKISAVGGYTDDDRPILMCVVGQTEFTK 251

Query: 247 LQEKILEIDDTAFIVVMPAAQVMGRGF 273
           L++ + +ID++AF++V  A++V+G GF
Sbjct: 252 LKQIVKQIDESAFVIVADASEVLGEGF 278
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 135/252 (53%), Positives = 190/252 (74%)

Query: 1   MAVSFHEVFGWDSAFFIMIINIPLLLLCYFGLGKQTFLKTVYGSWIFPVFIKLTQSVPTL 60
           +AV    +FG   + F+M  NIPLLL+CYF LGKQ F+KT+YGSWI+P+ I+ T S+PTL
Sbjct: 39  IAVVIKALFGISPSLFLMASNIPLLLMCYFFLGKQNFIKTLYGSWIYPIAIRSTNSLPTL 98

Query: 61  THNPLLAALFGGVIVGCGLGIVFWSDSSTGGTGIIIQFLGKYTPISLGQGVILIDGLVTI 120
           THN LLAA+FGG+I G GLG+VFW +SSTGGTGI+ Q L KY+P+SLG  + ++DG+  +
Sbjct: 99  THNQLLAAIFGGIICGIGLGMVFWGNSSTGGTGILTQILHKYSPLSLGVAMTIVDGISVL 158

Query: 121 VGFLAFDSDTVMFSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGV 180
           +GF+A  +D VM+S IGL  I Y+I+ ++ GF +   V+I+S+++Q I+ YI TV DRGV
Sbjct: 159 MGFIALSADDVMYSTIGLFVIGYVISVMENGFDSSKNVMIISKDYQAIREYITTVMDRGV 218

Query: 181 TEIPVKGGYSGTNQIMLMTTIAGYEFAKLQEAIAEIDETAFITVTPTSQASGRGFSLQKN 240
           T++P++GGY+ +++IMLM  ++ +E   LQE I EID+TAFI V P +Q  GRGFSL K
Sbjct: 219 TKLPIRGGYTTSDKIMLMAIVSSHELPTLQEKILEIDDTAFIVVMPAAQVMGRGFSLTKQ 278

Query: 241 HGRLDEDILMPM 252
           + R D+D+L+PM
Sbjct: 279 YKREDKDVLLPM 290
```

A related GBS gene <SEQ ID 8871> and protein <SEQ ID 8872> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:      1.57
GvH: Signal Score (-7.5): -2.56
     Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value:  -7.70 threshold:  0.0
   INTEGRAL    Likelihood = -7.70   Transmembrane   14 -  30 (   8 -  34)
   INTEGRAL    Likelihood = -6.90   Transmembrane   66 -  82 (  63 -  85)
   INTEGRAL    Likelihood = -6.69   Transmembrane  110 - 126 ( 105 - 128)
   INTEGRAL    Likelihood = -3.93   Transmembrane  132 - 148 ( 129 - 149)
   PERIPHERAL  Likelihood =  3.71      37
 modified ALOM score:   2.04

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02139(118 - 1008 of 1356)
OMNI|NT01BS4111(51 - 325 of 327) conserved hypothetical protein
%Match = 19.3
%Identity = 37.1  %Similarity = 62.1
Matches = 101  Mismatches = 99  Conservative Sub.s = 68

27        57        87       117                 165
           ARAIPSFIVGSALTGALVGLAGIKLMAPHGGIFVIALTSNPLLYIL----FILIGAVVSGVLFGLF---
                                                  |  |:||    :||||| :: | | :|
VCFFISYILDFTAALAYYHCIWVLFTSNQGRIKMLSESIGRNGGYMMDVRNKTLWILRDYVYILIGAAITAVSFNVFLLP
        10        20        30        40        50        60        70        80

216       246       276       306       336       366       396       426
RKIK*LISTYPNLH*IKGE*XIVILXXLIN*XXGGISGLAVSFXEVFGWDSAFFIMIINIPLLLLCYFGLGKQTFLKTVY
 ||                              ||:||::    : :|:::|:   ||||||::     || :  |||:
NKI----------------------------AAGGVSGIST-ILQSYGFEAAYVQWIINIPLFIAGVILLGGKFGLKTLA
                                    90         100       110       120       130

456       486       516       546       576       606       636       666
GSWIFPVFIKLTQSVPTLTHNPLLAALFGGVIVGCGLGIVFWSDSSTGGTGIIIQFLGKYTPISLGQGVILIDGLVTIVG
||  :|: : ||: :   ||: ||||:|||| :| |:|||:    ||||| :  | : ||: :|||: : :|||:: :
GSVFLPLVVFLTRDIQPATHHELLAAIFGGVGIGIGIGIVYLGKGSTGGTALAAQIIHKYSGLSLGKCLAIIDGMIVVTA
                 150       160       170       180       190       200       210

696       726       756       786       816       846       876       906
FLAFDSDTVMFSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGVTEIPVKGGYSGTNQIMLMTTIA
 : |: :  :::::|:   |  |: :| ||      ||:::: | :|  :    |||||:|   |||:  :: :||  :
MIVFNIEQGLYAMLGVYVSSKTIDVVQVGFNRSKMALIITKQEQAVKEAVLQKIDRGVTKISAVGGYTDDDRPILMCVVG
                 230       240       250       260       270       280       290

936       966       996      1026      1056      1086      1116
GYEFAKLQEAIAEIDETAFITVTPTSQASGRGFSLQKNHGRLDEDILMPM*SIDN*SFF**NSR*NIHKR*QNC
  || ||:: : :|||:||: |   |:  | ||
QTEFTKLKQIVKQIDESAFVIVADASEVLGEGFKRA
                 310       320
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1644

A DNA sequence (GBSx1739) was identified in *S.agalactiae* <SEQ ID 5087> which encodes the amino acid sequence <SEQ ID 5088>. This protein is predicted to be ABC transporter, ATP-binding protein (b0820). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3122(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24918 GB:AF012285 YkpA [Bacillus subtilis]
 Identities = 355/540 (65%), Positives = 451/540 (82%), Gaps = 4/540 (0%)

Query: 1    MLTVSDVSLRFSDRKLFDEVNINFTAGNTYGLIGANGAGKSTFLKILAGDIEPTTGHIAL 60
            M+ V++VSLRF+DRKLF++VNI FT GN YGLIGANGAGKSTFLK+L+G+IEP TG + +
Sbjct: 1    MIAVNNVSLRFADRKLFEDVNIKFTPGNCYGLIGANGAGKSTFLKVLSGEIEPQTGDVHM 60

Query: 61   GPDERLSVLRQNHFDYEDERVIDVVIMGNETLYSIMKEKDAIYMKEDFSDEDGVRAAELE 120
             P ERL+VL+QNHF+YE+  V+ VVIMG++ LY +M+EKDAIYMK DFSDEDG+RAAELE
Sbjct: 61   SPGERLAVLKQNHFEYEEYEVLKVVIMGHKRLYEVMQEKDAIYMKPDFSDEDGIRAAELE 120
```

```
Query: 121 GEFAELGGWEAESEASQLLQNLNISEELHYQNMSELANGDKVKVLLAKALFGKPDVLLLD 180
           GEFAEL GWEAESEA+ LL+ L ISE+LH + M++L   +KVKVLLA+ALFGKPDVLLLD
Sbjct: 121 GEFAELNGWEAESEAAILLKGLGISEDLHTKKMADLGGSEKVKVLLAQALFGKPDVLLLD 180

Query: 181 EPTNGLDIQSITWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW 240
           EPTN LD+Q+I WLE+FLI+FENTVIVVSHDRHFLNKVCTH+ADLDF KI+++VGNYDFW
Sbjct: 181 EPTNHLDLQAIQWLEEFLINFENTVIVVSHDRHFLNKVCTHIADLDFNKIQIYVGNYDFW 240

Query: 241 KESSELAARLQADRNAKAEEKIKQLQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS 300
            ESS+LA +L  + N K EE+IKQLQEFVARFSANASKSKQATSRKK+L+KI L++I PS
Sbjct: 241 YESSQLALKLSQEANKKKEEQIKQLQEFVARFSANASKSKQATSRKKLLEKITLDDIKPS 300

Query: 301 SRKYPFVNFKAEREMGNDLLTVENLSVTIDGEKILDNISFILRPGDKTALIGQNDIQTTA 360
           SR+YP+VNF  ERE+GND+L VE L+ TIDG K+LDN+SFI+   DK A  G+N++  T
Sbjct: 301 SRRYPYVNFTPEREIGNDVLRVEGLTKTIDGVKVLDNVSFIMNREDKIAFTGRNELAVTT 360

Query: 361 LIRALMGDIEYE-GTIKWGVTTSRSYLPKDNSRDFASGE-SILEWLRQFASKEEDDNTFL 418
           L + + G++E + GT KWGVTTS++Y PKDNS  F   + ++++WLRQ+ S  +   +FL
Sbjct: 361 LFKIISGEMEADSGTFKWGVTTSQAYFPKDNSEYFEGSDLNLVDWLRQY-SPHDQSESFL 419

Query: 419 RGFLGRMLFSGDEVNKSVNVLSGGEKVRVMLSKLMLLKSNVLVLDDPTNHLDLESISSLN 478
           RGFLGRMLFSG+EV+K  NVLSGGEKVR MLSK ML  +N+L+LD+PTNHLDLESI++LN
Sbjct: 420 RGFLGRMLFSGEEVHKKANVLSGGEKVRCMLSKAMLSGANILILDEPTNHLDLESITALN 479

Query: 479 DGLKDFKESIIFASHDHEFIQTLANHIIVLSKNGVIDRIDETYDEFLENTEVQAKVAQLW 538
           +GL  FK +++F SHDH+F+QT+AN II ++ NG++D+   +YDEFLEN +VQ K+ +L+
Sbjct: 480 NGLISFKGAMLFTSHDHQFVQTIANRIIEITPNGIVDK-QMSYDEFLENADVQKKLTELY 538
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5089> which encodes the amino acid sequence <SEQ ID 5090>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3124(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 497/539 (92%), Positives = 525/539 (97%)

Query: 1   MLTVSDVSLRFSDRKLFDEVNINFTAGNTYGLIGANGAGKSTFLKILAGDIEPTTGHIAL 60
           +LTVSDVSLRFSDRKLFD+VNI FTAGNTYGLIGANGAGKSTFLKILAGDIEP+TGHI+L
Sbjct: 1   LLTVSDVSLRFSDRKLFDDVNIKFTAGNTYGLIGANGAGKSTFLKILAGDIEPSTGHISL 60

Query: 61  GPDERLSVLRQNHFDYEDERVIDVVIMGNETLYSIMKEKDAIYMKEDFSDEDGVRAAELE 120
           GPDERLSVLRQNHFDYE+ER IDVVIMGNE LY+IMKEKDAIYMK DFS+EDGVRAAELE
Sbjct: 61  GPDERLSVLRQNHFDYEEERAIDVVIMGNEQLYNIMKEKDAIYMKADFSEEDGVRAAELE 120

Query: 121 GEFAELGGWEAESEASQLLQNLNISEELHYQNMSELANGDKVKVLLAKALFGKPDVLLLD 180
           G FAELGGWEAESEASQLLQNLNI E+LHYQNMSELANGDKVKVLLAKALFGKPDVLLLD
Sbjct: 121 GIFAELGGWEAESEASQLLQNLNIPEDLHYQNMSELANGDKVKVLLAKALFGKPDVLLLD 180

Query: 181 EPTNGLDIQSITWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW 240
           EPTNGLDIQSI+WLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW
Sbjct: 181 EPTNGLDIQSISWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW 240

Query: 241 KESSELAARLQADRNAKAEEKIKQLQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS 300
           K+SSELAARLQADRNAKAEEKIK+LQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS
Sbjct: 241 KQSSELAARLQADRNAKAEEKIKELQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS 300

Query: 301 SRKYPFVNFKAEREMGNDLLTVENLSVTIDGEKILDNISFILRPGDKTALIGQNDIQTTA 360
           SRKYPF+NFKAEREMGND LTVENLSVTIDGEKI+DNISFILRPGDK A+IGQNDIQTTA
Sbjct: 301 SRKYPFINFKAEREMGNDFLTVENLSVTIDGEKIIDNISFILRPGDKAAIIGQNDIQTTA 360
```

```
Query: 361 LIRALMGDIEYEGTIKWGVTTSRSYLPKDNSRDFASGESILEWLRQFASKEEDDNTFLRG 420
           L+RAL  DI+YEGTIKWGVTTSRSYLPKDNS+DFA+ ESILEWLRQFASK EDD+TFLRG
Sbjct: 361 LMRALADDIDYEGTIKWGVTTSRSYLPKDNSKDFATEESILEWLRQFASKGEDDDTFLRG 420

Query: 421 FLGRMLFSGDEVNKSVNVLSGGEKVRVMLSKLMLLKSNVLVLDDPTNHLDLESISSLNDG 480
           FLGRMLFSGDEV KSVNVLSGGEKVRVMLSKLMLLKSNVL+LDDPTNHLDLESISSLNDG
Sbjct: 421 FLGRMLFSGDEVKKSVNVLSGGEKVRVMLSKLMLLKSNVLILDDPTNHLDLESISSLNDG 480

Query: 481 LKDFKESIIFASHDHEFIQTLANHIIVLSKNGVIDRIDETYDEFLENTEVQAKVAQLWK 539
           +KDFKES+IFASHDHEFIQT+ANHI+V+SKNGVIDRIDETYDEFL+N EVQA+VA+LWK
Sbjct: 481 IKDFKESVIFASHDHEFIQTIANHIVVISKNGVIDRIDETYDEFLDNPEVQARVAELWK 539
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1645

A DNA sequence (GBSx1740) was identified in *S.agalactiae* <SEQ ID 5091> which encodes the amino acid sequence <SEQ ID 5092>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.71   Transmembrane   14 -  30 (   8 -  35)
   INTEGRAL    Likelihood = -7.70   Transmembrane  384 - 400 ( 382 - 403)
   INTEGRAL    Likelihood = -7.22   Transmembrane  412 - 428 ( 408 - 433)
   INTEGRAL    Likelihood = -5.73   Transmembrane  163 - 179 ( 155 - 180)
   INTEGRAL    Likelihood = -5.52   Transmembrane  322 - 338 ( 320 - 344)
   INTEGRAL    Likelihood = -5.10   Transmembrane  297 - 313 ( 290 - 314)
   INTEGRAL    Likelihood = -4.41   Transmembrane  360 - 376 ( 357 - 377)
   INTEGRAL    Likelihood = -4.35   Transmembrane  438 - 454 ( 437 - 455)
   INTEGRAL    Likelihood = -4.09   Transmembrane  136 - 152 ( 136 - 153)
   INTEGRAL    Likelihood = -3.35   Transmembrane  110 - 126 ( 106 - 128)
   INTEGRAL    Likelihood = -2.28   Transmembrane  232 - 248 ( 232 - 248)
   INTEGRAL    Likelihood = -1.81   Transmembrane  832 - 848 ( 832 - 848)
   INTEGRAL    Likelihood = -1.12   Transmembrane  200 - 216 ( 200 - 216)

----- Final Results -----
               bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC14608 GB:U95840 transmembrane protein Tmp5 [Lactococcus
          lactis]
 Identities = 140/260 (53%), Positives = 182/260 (69%), Gaps = 6/260 (2%)

Query: 16  SFLLPFIIIVCILFTKNIYWGSPTTILASDGFHQYVIFNQALRNILH--GSNSLFYTFTS 73
           SF +P I++V +L    IYWGS  +ILA D +HQYV  +   RNILH  GS    YTFTS
Sbjct: 14  SFFIPLILMVIVLAMTGIYWGSSRSILAGDAYHQYVAIHSLYRNILHSGGSQGFLYTFTS 73

Query: 74  GLGLNFYALSSYYLGSFLSPIVYFFNLKNMPDAIYLLTICKIGLIGLSMFVTLCKRHCKV 133
           GLGLN YA S+YY+GSFL P  +FF++K+MPDA+YL TI K GLIGLS FV+    + K+
Sbjct: 74  GLGLNLYAFSAYYMGSFLMPFTFFFDVKSMPDALYLFTIIKFGLIGLSSFVSFKNMYQKL 133

Query: 134 NRVLLLVISTCYSLMSFSISQIEINMWLDVFILIPLVVLGVDQLLWERKPILYFLSLTAL 193
           + + +L IST ++LMSF  SQ+EI MWLDVFIL+PL++ G+ +L+ ERK  LYF+SL  L
Sbjct: 134 SNLTVLSISTAFALMSFLTSQLEITMWLDVFILLPLIIWGLHRLMDERKRWLYFVSLLIL 193

Query: 194 FIQNYYFGFMTAIFTSLYFIVQITRNTDSKVAFKQFLHFTFLSLLAGMTSSIMILPTYFD 253
           FIQNYYFGFM AIF  LYF   + R T  K ++ + L F   S LAG+ S IM+LP Y D
Sbjct: 194 FIQNYYFGFMVAIFLVLYF---LARMTYEKWSWTKVLDFVVSSTLAGIASLIMLLPMYLD 250

Query: 254 L-TTHGEKLTKVSKMFTENS 272
           L + + + L+ +S +FTENS
```

```
Sbjct: 251 LKSNNSDALSTLSGIFTENS 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5093> which encodes the amino acid sequence <SEQ ID 5094>. Analysis of this protein sequence reveals the following:

```
        Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.29   Transmembrane   15 -  31 (   6 -  35)
   INTEGRAL    Likelihood = -8.81   Transmembrane  201 - 217 ( 196 - 220)
   INTEGRAL    Likelihood = -6.79   Transmembrane  410 - 426 ( 402 - 428)
   INTEGRAL    Likelihood = -6.05   Transmembrane  230 - 246 ( 227 - 252)
   INTEGRAL    Likelihood = -5.57   Transmembrane  161 - 177 ( 153 - 178)
   INTEGRAL    Likelihood = -4.46   Transmembrane  291 - 307 ( 290 - 311)
   INTEGRAL    Likelihood = -3.82   Transmembrane  133 - 149 ( 130 - 151)
   INTEGRAL    Likelihood = -3.77   Transmembrane  380 - 396 ( 376 - 400)
   INTEGRAL    Likelihood = -3.61   Transmembrane  105 - 121 ( 103 - 124)
   INTEGRAL    Likelihood = -3.45   Transmembrane  832 - 848 ( 830 - 848)
   INTEGRAL    Likelihood = -2.66   Transmembrane  436 - 452 ( 435 - 453)
   INTEGRAL    Likelihood = -2.13   Transmembrane  318 - 334 ( 314 - 336)
   INTEGRAL    Likelihood = -1.54   Transmembrane  356 - 372 ( 355 - 372)
   INTEGRAL    Likelihood = -0.27   Transmembrane   80 -  96 (  80 -  96)

----- Final Results -----
              bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC14608 GB:U95840 transmembrane protein Tmp5 [Lactococcus lactis]
 Identities = 134/269 (49%), Positives = 183/269 (67%), Gaps = 8/269 (2%)

Query: 5   NKWIIAGLASFLFPLSIIFIILLSMGIYYNSDKTILASDAFHQYVIFAQNFRNIMH--GS 62
            NKW +  LASF  PL ++ I+L   GIY+ S ++ILA DA+HQYV     +RNI+H  GS
Sbjct: 7   NKWAL--LASFFIPLILMVIVLAMTGIYWGSSRSILAGDAYHQYVAIHSLYRNILHSGGS 64

Query: 63  DSFFYTFTSGLGINFYALMCYYLGSFFSPLLFFFNLTSMPDAIYLFTLIKFGLIGLAACY 122
              F YTFTSGLG+N YA   YY+GSF  P  FFF++ SMPDA+YLFT+IKFGLIGL++
Sbjct: 65  QGFLYTFTSGLGLNLYAFSAYYMGSFLMPFTFFFDVKSMPDALYLFTIIKFGLIGLSSFV 124

Query: 123 SFHRLYPKISAFLMISISVFYSLMSFLTSQMELNSWLDVFILLPLVILGLNKLITENKTR 182
            SF  +Y K+S   ++SIS  ++LMSFLTSQ+E+  WLDVFILLPL+I GL++L+ E K
Sbjct: 125 SFKNMYQKLSNLTVLSISTAFALMSFLTSQLEITMWLDVFILLPLIIWGLHRLMDERKRW 184

Query: 183 TYYLSISLLFIQNYYFGYMIALFCILYALVCLLRLNDFNKMFIAFVRFTAVSICAALTSA 242
             Y++S+ +LFIQNYYFG+M+A+F +LY L    R+      +   + F   S  A + S
Sbjct: 185 LYFVSLLILFIQNYYFGFMVAIFLVLYFLA---RMTYEKWSWTKVLDFVVSSTLAGIASL 241

Query: 243 LVILPTYLDL-STYGENLSPIKQLVTNNA 270
            +++LP YLDL S   + LS +  + T N+
Sbjct: 242 IMLLPMYLDLKSNNSDALSTLSGIFTENS 270
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 432/836 (51%), Positives = 569/836 (67%), Gaps = 2/836 (0%)

Query: 16  SFLLPFIIIVCILFTKNIYWGSPTTILASDGFHQYVIFNQALRNILHGSNSLFYTFTSGL 75
            SFL P  II  IL +  IY+ S  TILASD FHQYVIF Q  RNI+HGS+S FYTFTSGL
Sbjct: 14  SFLFPLSIIFIILLSMGIYYNSDKTILASDAFHQYVIFAQNFRNIMHGSDSFFYTFTSGL 73

Query: 76  GLNFYALSSYYLGSFLSPIVYFFNLKNMPDAIYLLTICKIGLIGLSMFVTLCKRHCKVNR 135
            G+NFYAL  YYLGSF SP+++FFNL +MPDAIYL T+ K GLIGL+   +  + + K++
Sbjct: 74  GINFYALMCYYLGSFFSPLLFFFNLTSMPDAIYLFTLIKFGLIGLAACYSFHRLYPKISA 133

Query: 136 VLLLVISTCYSLMSFSISQIEINMWLDVFILIPLVVLGVDQLLWERKPILYFLSLTALFI 195
             L++ IS  YSLMSF  SQ+E+N WLDVFIL+PLV+LG+++L+ E K   Y+LS++ LFI
Sbjct: 134 FLMISISVFYSLMSFLTSQMELNSWLDVFILLPLVILGLNKLITENKTRTYYLSISLLFI 193
```

```
Query: 196 QNYYFGFMTAIFTSLYFIVQITRNTDSKVAFKQFLHFTFLSLLAGMTSSIMILPTYFDLT 255
           QNYYFG+M A+F  LY +V + R  D    F  F+ FT +S+ A +TS+++ILPTY DL+
Sbjct: 194 QNYYFGYMIALFCILYALVCLLRLNDFNKMFIAFVRFTAVSICAALTSALVILPTYLDLS 253

Query: 256 THGEKLTKVSKMFTENSWYMDLFAKNMIGAYDTTKFGSIPMIYVGLLPLLLSLLYFTIKE 315
           T+GE L+ + ++ T N+W++D+ AK  IG YDTTKF ++PMIYVGL PL+LS++YFT++
Sbjct: 254 TYGENLSPIKQLVTNNAWFLDIPAKLSIGVYDTTKFNALPMIYVGLFPLMLSVIYFTLES 313

Query: 316 VPRRTRLAYGFLIIFVIASFYITPLDLFWQGMHAPNMFLHRYSWVLSVLICLLAAECLEY 375
           +P + +LA   L+ F+I SFY+ PLDLFWQGMH+PNMFLHRY+W  S++I LLA E L
Sbjct: 314 IPLKIKLANACLLTFIIISFYLQPLDLFWQGMHSPNMFLHRYAWSFSIVILLLACETLSR 373

Query: 376 LDNISWKKILGVNLILVSGFIITFLFKKHYHYLNLELLLLTLTFLSAYIILTISFVSKQI 435
           L ++   K     + L+    + + F + Y++L L L LL++  L  Y I   SF + QI
Sbjct: 374 LKEVTQIKAGFAFIFLIILTSLPYSFSQQYNFLPLTLFLLSVFLLLGYTISLFSFRNSQI 433

Query: 436 PKLVFYPFLIGFVVLEMTLNTFYQLNSLNDEWIFPSRQGYAKYNHSISKLVRKTERNNST 495
           P      F++ F +LE  LNT+YQL  +N EW FPSRQ Y      I+ LV    +N+
Sbjct: 434 PSTFISAFILIFSLLESGLNTYYQLQGINKEWGFPSRQIYNSQLKDINNLVNSVSKNSQP 493

Query: 496 FFRTERWLGQTGNDSMKYNYNGISQFSSIRNRSSSQVLDRLGFKSDGTNLNLRYQNNTLI 555
           FFR ER L QTGNDSMK+NY GISQFSS+RNR SS +LDRLGF+S GTNLNLRYQNNT+I
Sbjct: 494 FFRMERLLPQTGNDSMKFNYYGISQFSSVRNRLSSSLLDRLGFQSKGTNLNLRYQNNTII 553

Query: 556 ADSLFGVKYNLTEYPFDKFGFIKKAQDKQTILYKNQFASQLAILTNQVYQDKPFTVNTLD 615
            DSL G+KYNL+E P +KFGF K      T LY+N ++S LAILT  VY+D    VNTLD
Sbjct: 554 MDSLLGIKYNLSEGPPNKFGFTKLKTSGNTTLYQNHYSSPLAILTRNVYKDVNLNVNTLD 613

Query: 616 NQTTLLNQLSGLKETYFEHLIPNSVSGQTTLNKQVFVK-KNKQGNTEITYNITIPKNSQL 674
           NQT LLNQLSG   TYF       +SG    N Q+  +  + Q +  + Y I IPK+SQL
Sbjct: 614 NQTKLLNQLSGKSLTYFNLQPAQLISGANQFNGQISAQASDYQNSVTLNYQINIPKHSQL 673

Query: 675 YVSMPFINFNNEENKIVQISVNNGPFVPNTLDNAYSFFNIGSFAENSRIKVKFQFPHNDQ 734
           YVS+P I F+N + K ++I  +N  F+  T DNAYSFF++G FA+       F FP N Q
Sbjct: 674 YVSIPNIIFSNPDAKEMRIQTDNHNFI-YTTDNAYSFFDLGYFADAKVATFSFVFPKNKQ 732

Query: 735 VSFPIPHFYGLKLEAYQKAMTVINKRKVKVRTDHNKVIANYTSPNRSSLFFTIPYDRGWK 794
           +SF  PHFY L +E+Y +AM  I ++ V      N VI +Y S  + SL FT+PYD+GW
Sbjct: 733 ISFKEPHFYSLSIESYLEAMNSIKQKNVHTYAKSNTVITDYNSKTKGSLIFTLPYDKGWS 792

Query: 795 AYQNNKEIKIFKAQKGFMKINIPKGKGKVTLIFIPYGFKFGVGLSITGIVLFTVYY 850
           A ++ K + + KAQ GF+ + IPKGKG+V L FIP GFK G+ LS  GI+ + + Y
Sbjct: 793 AQKDGKNLPVKKAQGGFLSVTIPKGKGRVILTFIPNGFKLGLSLSCVGIIAYMLLY 848
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1646

A DNA sequence (GBSx1741) was identified in *S.agalactiae* <SEQ ID 5095> which encodes the amino acid sequence <SEQ ID 5096>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4624(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45340 GB:AF000658 ORF1 [Streptococcus pneumoniae]
 Identities = 111/159 (69%), Positives = 136/159 (84%)
```

```
Query: 1   MKLKIITVGKLKEKYLKEGVAEYQKRLNRFSKIETIELADEKTPDKASISENQRILDIEG 60
           MK+K++TVGKLKEKYLK+G+AEY KR++RF+K E IEL+DEKTPDKAS SENQ+IL+IEG
Sbjct: 1   MKIKVVTVGKLKEKYLKDGIAEYSKRISRFAKFEMIELSDEKTPDKASESENQKILEIEG 60

Query: 61  ERILSKIGERDYVIGLAIEGKQLPSESFSHLIDQKMISGYSTITFVIGGSLGLSQKVKKR 120
           +RILSKI +RD+VI LAIEGK   SE FS  +++  I G+ST+TF+IGGSLGLS  VK R
Sbjct: 61  QRILSKIADRDFVIVLAIEGKTFFSEEFSKQLEETSIKGFSTLTFIIGGSLGLSSSVKNR 120

Query: 121 ADYLMSFGLLTLPHQLMKLVLMEQIYRAFMIRQGTPYHK 159
           A+  +SFG LTLPHQLM+LVL+EQIYRAF I+QG PYHK
Sbjct: 121 ANLSVSFGRLTLPHQLMRLVLVEQIYRAFTIQQGFPYHK 159
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5097> which encodes the amino acid sequence <SEQ ID 5098>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 112/159 (70%), Positives = 133/159 (83%)

Query: 1   MKLKIITVGKLKEKYLKEGVAEYQKRLNRFSKIETIELADEKTPDKASISENQRILDIEG 60
           MK+K+I VGKLKE+YLK+G++EYQKRL+RF + E IEL DE+TPDKAS ++NQ I+  E
Sbjct: 1   MKVKLICVGKLKERYLKDGISEYQKRLSRFCQFEMIELTDERTPDKASFADNQLIMSKEA 60

Query: 61  ERILSKIGERDYVIGLAIEGKQLPSESFSHLIDQKMISGYSTITFVIGGSLGLSQKVKKR 120
           +RI  KIGERD+VI LAIEGKQ PSE+FS LI    + GYSTITF+IGGSLGL   +KKR
Sbjct: 61  QRIHKKIGERDFVIALAIEGKQFPSETFSELISGVTVKGYSTITFIIGGSLGLDSIIKKR 120

Query: 121 ADYLMSFGLLTLPHQLMKLVLMEQIYRAFMIRQGTPYHK 159
           A+ LMSFGLLTLPHQLM+LVL EQIYRAFMI QG+PYHK
Sbjct: 121 ANMLMSFGLLTLPHQLMRLVLTEQIYRAFMITQGSPYHK 159
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1647

A DNA sequence (GBSx1742) was identified in *S.agalactiae* <SEQ ID 5099> which encodes the amino acid sequence <SEQ ID 5100>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3785(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1648

A DNA sequence (GBSx1743) was identified in *S.agalactiae* <SEQ ID 5101> which encodes the amino acid sequence <SEQ ID 5102>. This protein is predicted to be a serine protease. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4533(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9445> which encodes amino acid sequence <SEQ ID 9446> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45334 GB:AF000658 putative serine protease [Streptococcus pneumoniae]
 Identities = 215/370 (58%), Positives = 278/370 (75%), Gaps = 20/370 (5%)

Query: 4   NDNIPNGGVTKTSKVNYNNITPTTKAVKKVQNSVVSVINYKQQESRSDLSDFYSHFFGNQ 63
           N++  N  +T+T+   Y N   TT+AV KV+++VVSVI Y      S         FGN
Sbjct: 46  NNSNNNSTITQTA---YKNENSTTQAVNKVKDAVVSVITYSANRQNS--------VFGND 94

Query: 64  GGNTDKGLQVYGEGSGVIYKKDGKNAYVVTNNHVIDGAKQIEIQLADGSKAVGKLVGSDT 123
             +TD   ++  EGSGVIYKK+ K AY+VTNNHVI+GA +++I+L+DG+K  G++VG+DT
Sbjct: 95  DTDTDSQ-RISSEGSGVIYKKNDKEAYIVTNNHVINGASKVDIRLSDGTKVPGEIVGADT 153

Query: 124 YSDLAVVKIPSDKVSNIAEFADSSKLNIGETAIAIGSPLGTEYANSVTQGIVSSLKRTVT 183
           +SD+AVVKI S+KV+ +AEF DSSKL +GETAIAIGSPLG+EYAN+VTQGIVSSL R V+
Sbjct: 154 FSDIAVVKISSEKVTTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQGIVSSLNRNVS 213

Query: 184 MTNEEGQTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSGQSSGNSV 243
           + +E+GQ +ST AIQTD AINPGNSGG LINI+GQVIGI SSKI++        + G SV
Sbjct: 214 LKSEDGQAISTKAIQTDTAINPGNSGGPLINIQGQVIGITSSKIAT--------NGGTSV 265

Query: 244 EGMGFAIPSNDVVKIINQLESNGQVERPALGISMAGLSNLPSDVISKLKIPSNVTNGIVV 303
           EG+GFAIP+ND + II QLE NG+V RPALGI M  LSN+ +  I +L IPSNVT+G++V
Sbjct: 266 EGLGFAIPANDAINIIEQLEKNGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSGVIV 325

Query: 304 ASIQSGMPAQGKLKKYDVITKVDDKEVVSPSDLQSLLYGHQVGDSITVTFYRGENKQTVT 363
            S+QS MPA G L+KYDVITKVDDKE+ S +DLQS LY H +GD+I +T+YR   ++T +
Sbjct: 326 RSVQSNMPANGHLEKYDVITKVDDKEIASSTDLQSALYNHSIGDTIKITYYRNGKEETTS 385

Query: 364 IKLTKTSKDL 373
           IKL K+S DL
Sbjct: 386 IKLNKSSGDL 395
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5103> which encodes the amino acid sequence <SEQ ID 5104>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.76    Transmembrane    11 -  27 (    6 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 250/375 (66%), Positives = 299/375 (79%), Gaps = 5/375 (1%)
```

```
Query: 3   HNDNIPNGGVTKTSKVNYNNITPTTKAVKKVQNSVVSVINYKQQESRSDLSDFYSHFFGN 62
           H+ +  N G   TS + +NN T TTKAVK VQN+VVSVINY+   S S LS+ Y+  FG
Sbjct: 34  HSPSKINSGKATTSNMVFNNTTNTTKAVKAVQNAVVSVINYQDNPS-SSLSNPYTKLFGE 92

Query: 63  QGG--NTDKGLQVYGEGSGVIYKKDGKNAYVVTNNHVIDGAKQIEIQLADGSKAVGKLVG 120
                N D  L ++ EGSGVIY+KDG +AYVVTNNHVIDGAK+IEI +ADGSK VG+LVG
Sbjct: 93  GRSKENKDAELSIFSEGSGVIYRKDGNSAYVVTNNHVIDGAKRIEILMADGSKVVGELVG 152

Query: 121 SDTYSDLAVVKIPSDKVSNIAEFADSSKLNIGETAIAIGSPLGTEYANSVTQGIVSSLKR 180
           +DTYSDLAVVKI SDK+  +AEFADS+KLN+GE AIAIGSPLGT+YANSVTQGIVSSL R
Sbjct: 153 ADTYSDLAVVKISSDKIKTVAEFADSTKLNVGEVAIAIGSPLGTQYANSVTQGIVSSLSR 212

Query: 181 TVTMTNEEGQTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSGQSSG 240
           TVT+ NE G+TVSTNAIQTDAAINPGNSGG LINIEGQVIGINSSKISST   ++G S
Sbjct: 213 TVTLKNENGETVSTNAIQTDAAINPGNSGGPLINIEGQVIGINSSKISSTPTGSNGNS-- 270

Query: 241 NSVEGMGFAIPSNDVVKIINQLESNGQVERPALGISMAGLSNLPSDVISKLKIPSNVTNG 300
            +VEG+GFAIPS DV+KII QLE+NG+V RPALGISM  L++L ++ +S++ IP++VT G
Sbjct: 271 GAVEGIGFAIPSTDVIKIIKQLETNGEVIRPALGISMVNLNDLSTNALSQINIPTSVTGG 330

Query: 301 IVVASIQSGMPAQGKLKKYDVITKVDDKEVVSPSDLQSLLYGHQVGDSITVTFYRGENKQ 360
           IVVA ++ GMPA GKL +YDVIT++D K V S SDLQS LYGH + D+I VTFYRG  K+
Sbjct: 331 IVVAEVKEGMPASGKLAQYDVITEIDGKTVNSISDLQSSLYGHDINDTIKVTFYRGTTKK 390

Query: 361 TVTIKLTKTSKDLAK 375
              IKLTKT++DL K
Sbjct: 391 KADIKLTKTTQDLTK 405
```

A related GBS gene <SEQ ID 8873> and protein <SEQ ID 8874> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     12.68
GvH: Signal Score (-7.5): -1.33
     Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:    4.56 threshold:  0.0
   PERIPHERAL  Likelihood =  4.56     301
 modified ALOM score:  -1.41

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.4/75.6% over 386aa
                                                                   Streptococcus
pneumoniae
  GP|2109443| putative serine protease Insert characterized

ORF02135(307 - 1506 of 1827)
GP|2109443|gb|AAC45334.1||AF000658(9 - 395 of 397) putative serine protease {Streptococcus
pneumoniae}
%Match = 34.6
%Identity = 57.3  %Similarity = 75.6
Matches = 223  Mismatches = 89  Conservative Sub.s = 71

228       258       288       318       348       378       399       429
RLSTSCGYFLFLAFKV*LRSLS*D*YKNLRR*LFVKKKLVSSLLKCSLIIIVSFAGGAFASFVMNH---NDNIPNGGVTK
                            :   :   | :    ::  ::|::||  ||: || : :     :: |
                            MEANMKHLKTFYKKWFQLLVVIVISFFSGALGSFSITQLTQKSSVNNSNNNS
                                     10        20        30        40        50
```

```
456       486       516       546       576       606       636       666
T-SKVNYNNITPTTKAVKKVQNSVVSVINYKQQESRSDLSDFYSHFFGNQGGNTDKGLQVYGEGSGVIYKKDGKNAYVVT
| ::  | |   ||:|| ||:::||||| |      |         |||   :||   ::  |||||||||: | ||:||
TITQTAYKNENSTTQAVNKVKDAVVSVITYSANRQNS--------VFGNDDTDTDS-QRISSEGSGVIYKKNDKEAYIVT
                70        80                90       100        110       120

696       726       756       786       816       846       876       906
NNHVIDGAKQIEIQLADGSKAVGKLVGSDTYSDLAVVKIPSDKVSNIAEFADSSKLNIGETAIAIGSPLGTEYANSVTQG
|||||:|| :::|:|:||:|  |::||:||:||:||||| |:||: :||| ||||| :||||||||||||:||||:||||
NNHVINGASKVDIRLSDGTKVPGEIVGADTFSDIAVVKISSEKVTTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQG
              140       150       160       170       180       190       200

936       966       996      1026      1056      1086      1116      1146
IVSSLKRTVTMTNEEGQTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSGQSSGNSVEGMGFAIPSN
||||| | |:: :|:|| :|| ||||| ||||||||| ||||:|||||| ||||:     |:|   | ||||:|||||:|
IVSSLNRNVSLKSEDGQAISTKAIQTDTAINPGNSGGPLINIQGQVIGITSSKIA-----TNG---GTSVEGLGFAIPAN
              220       230       240       250            260          270

1176      1206      1236      1266      1296      1326      1356      1386
DVVKIINQLESNGQVERPALGISMAGLSNLPSDVISKLKIPSNVTNGIVVASIQSGMPAQGKLKKYDVITKVDDKEVVSP
| : || ||| ||:| |||||| |  |||: :  | :| ||||||:|::| |:|| ||| | |:||||||||||||: |
DAINIIEQLEKNGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSGVIVRSVQSNMPANGHLEKYDVITKVDDKEIASS
            290       300       310       320       330       340       350

1416      1446      1476      1506      1536      1566      1596      1626
SDLQSLLYGHQVGDSITVTFYRGENKQTVTIKLTKTSKDLAKQRANN*INSSYFN*DIVKLKGLVR*TNPFSKSIESEV*
:|||| || | :||:| :|:||   ::| :||| |:| ||
TDLQSALYNHSIGDTIKITYYRNGKEETTSIKLNKSSGDLES
            370       380       390
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1649

A DNA sequence (GBSx1744) was identified in *S.agalactiae* <SEQ ID 5105> which encodes the amino acid sequence <SEQ ID 5106>. This protein is predicted to be SPSpoJ (spo0J). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4152(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45335 GB:AF000658 SPSpoJ [Streptococcus pneumoniae]
 Identities = 138/257 (53%), Positives = 188/257 (72%), Gaps = 5/257 (1%)

Query: 1   MEYLETININHIAPNPYQPRLEFNTKELEELANSIKINGLIQPIIVRPSAVFGYELVAGE 60
           ME  E I+I  I  NPYQPR EF+ ++L+ELA SIK NG+IQPIIVR S V GYE++AGE
Sbjct: 1   MEKFEMISITDIQKNPYQPRKEFDREKLDELAQSIKENGVIQPIIVRQSPVIGYEILAGE 60

Query: 61  RRLRAAKLAKLESIPAIIKSYNNDDSMQLAIVENLQRSNLSPIEEAKAYSQLLQKKSMTH 120
           RR RA+ LA L SIPA++K  ++ + M  +I+ENLQR NL+PIEEA+AY  L++ K  TH
Sbjct: 61  RRYRASLLAGLRSIPAVVKQISDQEMMVQSIIENLQRENLNPIEEARAYVSLVE-KGFTH 119

Query: 121 EELAKYMGKSRPYISNTIRLLNLPPLITSAIEEGKLSSGHARALLSLPDASQQKDWYQRI 180
            E+A   GKSRPYISN+IRLL+LP  I S +E GKLS  HAR+L+ L +  QQ  ++QRI
Sbjct: 120 AEIADKEGKSRPYISNSIRLLSLPEQILSEVENGKLSQAHARSLVGL-NKEQQDYFFQRI 178
```

```
Query: 181 LTEDISVRRLEKLLKQEKKTNHKSLQNKDVFLKHQENELAQFLGSKVKLTINKDGAGNIK 240
           + EDISVR+LE LL ++K+   K  Q  + F++++E +L + LG  V++ ++K  +G I
Sbjct: 179 IEEDISVRKLEALLTEKKQ---KKQQKTNHFIQNEEKQLRKLLGLDVEIKLSKKDSGKII 235

Query: 241 IAFANQEELNRIINTLK 257
           I+F+NQEE +RIIN+LK
Sbjct: 236 ISFSNQEEYSRIINSLK 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5107> which encodes the amino acid sequence <SEQ ID 5108>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1758(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/256 (57%), Positives = 191/256 (74%), Gaps = 1/256 (0%)

Query: 2   EYLETININHIAPNPYQPRLEFNTKELEELANSIKINGLIQPIIVRPSAVFGYELVAGER 61
           E L  + I  I  NPYQPR++FN +EL++LA SIK NGLIQPIIVR S +FGYELVAGER
Sbjct: 14  ELLIDLPIEDIVTNPYQPRIQFNQRELQDLATSIKSNGLIQPIIVRKSDIFGYELVAGER 73

Query: 62  RLRAAKLAKLESIPAIIKSYNNDDSMQLAIVENLQRSNLSPIEEAKAYSQLLQKKSMTHE 121
           RL+A+K+A L+ +PAIIK  +  +SMQ AIVENLQRSNL+ IEEAKAY  L++KK MTH+
Sbjct: 74  RLKASKMAGLKKVPAIIKKISTLESMQQAIVENLQRSNLNAIEEAKAYQLLVEKKHMTHD 133

Query: 122 ELAKYMGKSRPYISNTIRLLNLPPLITSAIEEGKLSSGHARALLSLPDASQQKDWYQRIL 181
           E+AKYMGKSRPYISNT+RLL LP  I  AIEEGK+S+GHARALL+L D  QQ     +I
Sbjct: 134 EIAKYMGKSRPYISNTLRLLQLPAPIIKAIEEGKISAGHARALLTLSDDKQQLYLTHKIQ 193

Query: 182 TEDISVRRLEKLLKQEKKTNHKSLQNKDVFLKHQENELAQFLGSKVKLTINKDGAGNIKI 241
            E +SVR++E+L+     ++  S + K++F    E +LA+ LG  V + +  + +G ++I
Sbjct: 194 NEGLSVRQIEQLV-TSTPSSKLSKKTKNIFATSLEKQLAKSLGLSVNMKLTANHSGYLQI 252

Query: 242 AFANQEELNRIINTLK 257
           +F+N +ELNRIIN LK
Sbjct: 253 SFSNDDELNRIINKLK 268
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1650

A DNA sequence (GBSx1745) was identified in *S.agalactiae* <SEQ ID 5109> which encodes the amino acid sequence <SEQ ID 5110>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10297> which encodes amino acid sequence <SEQ ID 10298> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5111> which encodes the amino acid sequence <SEQ ID 5112>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3646(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 353/455 (77%), Positives = 401/455 (87%), Gaps = 6/455 (1%)

Query: 32  MTENEQLFWNRVLELSRSQIAPAAYEFFVLEARLLKIEHQTAVITLDNIEMKKLFWEQNL 91
            MTENEQ+FWNRVLEL++SQ+   A YEFFV +ARLLK++   A I LD  +MK+LFWE+NL
Sbjct: 1   MTENEQIFWNRVLELAQSQLKQATYEFFVHDARLLKVDKHIATIYLD--QMKELFWEKNL 58

Query: 92  GPVILTAGFEIFNAEITANYV-SNDLHLQETSFS-NYQQSSNEVNTLPIRKIDSNLKEKY 149
              VILTAGFE++NA+I+ +YV   DL +++     N +     +N+LP   + S+L  KY
Sbjct: 59  KDVILTAGFEVYNAQISVDYVFEEDLMIEQNQTKINQKPKQQALNSLPT--VTSDLNSKY 116

Query: 150 TFANFVQGDENRWAVSASIAVADSPGTTYNPLFIWGGPGLGKTHLLNAIGNQVLRDNPNA 209
            +F NF+QGDENRWAV+ASIAVA++PGTTYNPLFIWGGPGLGKTHLLNAIGN VL +NPNA
Sbjct: 117 SFENFIQGDENRWAVAASIAVANTPGTTYNPLFIWGGPGLGKTHLLNAIGNSVLLENPNA 176

Query: 210 RVLYITAENFINEFVSHIRLDSMEELKEKFRNLDLLLIDDIQSLAKKTLGGTQEEFFNTF 269
            R+ YITAENFINEFV HIRLD+M+ELKEKFRNLDLLLIDDIQSLAKKTL GTQEEFFNTF
Sbjct: 177 RIKYITAENFINEFVIHIRLDTMDELKEKFRNLDLLLIDDIQSLAKKTLSGTQEEFFNTF 236

Query: 270 NALHTNDKQIVLTSDRNPNQLNDLEERLVTRFSWGLPVNITPPDFETRVAILTNKIQEYP 329
            NALH N+KQIVLTSDR P+ LNDLE+RLVTRF WGL VNITPPDFETRVAILTNKIQEY
Sbjct: 237 NALHNNNKQIVLTSDRTPDHLNDLEDRLVTRFKWGLTVNITPPDFETRVAILTNKIQEYN 296

Query: 330 YDFPQDTIEYLAGEFDSNVRELEGALKNISLVADFKHAKTITVDIAAEAIRARKNDGPIV 389
            + FPQDTIEYLAG+FDSNVR+LEGALK+ISLVA+FK   TITVDIAAEAIRARK DGP +
Sbjct: 297 FIFPQDTIEYLAGQFDSNVRDLEGALKDISLVANFKQIDTITVDIAAEAIRARKQDGPKM 356

Query: 390 TVIPIEEIQIQVGKFYGVTVKEIKATKRTQDIVLARQVAMYLAREMTDNSLPKIGKEFGG 449
            TVIPIEEIQ QVGKFYGVTVKEIKATKRTQ+IVLARQVAM+LAREMTDNSLPKIGKEFGG
Sbjct: 357 TVIPIEEIQAQVGKFYGVTVKEIKATKRTQNIVLARQVAMFLAREMTDNSLPKIGKEFGG 416

Query: 450 RDHSTVLHAYNKIKNMVAQDDNLRIEIETIKNKIR 484
            RDHSTVLHAYNKIKNM++QD++LRIEIETIKNKI+
Sbjct: 417 RDHSTVLHAYNKIKNMISQDESLRIEIETIKNKIK 451
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1651

A DNA sequence (GBSx1746) was identified in *S.agalactiae* <SEQ ID 5113> which encodes the amino acid sequence <SEQ ID 5114>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0556(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45337 GB:AF000658 beta subunit of DNA polymerase III
           [Streptococcus pneumoniae]
 Identities = 278/378 (73%), Positives = 324/378 (85%)

Query: 1   MIHFSINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAIILTGSNGQISIENTIPA 60
            MIHFSINKN FL AL +TKRAIS KNAIPILSTVKI+VT + + L GSNGQISIEN I
Sbjct: 1   MIHFSINKNLFLQALNITKRAISSKNAIPILSTVKIDVTNEGVTLIGSNGQISIENFISQ 60

Query: 61  SNENAGLLVTNPGSILLEAGFFINIISSLPDVTLEFTEIEQHQIVLTSGKSEITLKGKDV 120
             NE+AGLL+T+ GSILLEA FFIN++SSLPDVTL+F EIEQ+QIVLTSGKSEITLKGKD
Sbjct: 61  KNEDAGLLITSLGSILLEASFFINVVSSLPDVTLDFKEIEQNQIVLTSGKSEITLKGKDS 120

Query: 121 DQYPRLQEMTTDTPLTLETKLLKSIINETAFAASQQESRPILTGVHLVISQNKYFKAVAT 180
            +QYPR+QE++  TPL LETKLLK IINETAFAAS QESRPILTGVH V+SQ+K  K VAT
Sbjct: 121 EQYPRIQEISASTPLILETKLLKKIINETAFAASTQESRPILTGVHFVLSQHKELKTVAT 180

Query: 181 DSHRMSQRTFQLEKSANNFDLVVPSKSLREFSAVFTDDIETVEVFFSDSQMLFRSENISF 240
            DSHR+SQ+   LEK++++FD+V+PS+SLREFSAVFTDDIETVE+FF+++Q+LFRSENISF
Sbjct: 181 DSHRLSQKKLTLEKNSDDFDVVIPSRSLREFSAVFTDDIETVEIFFANNQILFRSENISF 240

Query: 241 YTRLLEGNYPDTDRLLTNQFETEIIFNTNALRHAMERAYLISNATQNGTVRLEIQNETVS 300
            YTRLLEGNYPDTDRL+   F T I FN   LR +MERA L+S+ATQNGTV+LEI++  VS
Sbjct: 241 YTRLLEGNYPDTDRLIPTDFNTTITFNVVNLRQSMERARLLSSATQNGTVKLEIKDGVVS 300

Query: 301 AHVNSPEVGKVNEELDTVSLKGDSLNISFNPTYLIESLKAVKSETVTIRFISPVRPFTLT 360
            AHV+SPEVGKVNEE+DT  + G+ L ISFNPTYLI+SLKA+ SE VTI FIS VRPFTL
Sbjct: 301 AHVHSPEVGKVNEEIDTDQVTGEDLTISFNPTYLIDSLKALNSEKVTISFISAVRPFTLV 360

Query: 361 PGEDTEDFIQLITPVRTN 378
            P +  EDF+QLITPVRTN
Sbjct: 361 PADTDEDFMQLITPVRTN 378
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5115> which encodes the amino acid sequence <SEQ ID 5116>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70    Transmembrane    67 -  83 (  67 -  83)

----- Final Results -----
            bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 295/378 (78%), Positives = 334/378 (88%)

Query: 1   MIHFSINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAIILTGSNGQISIENTIPA 60
            MI FSIN+  F+HAL  TKRAIS KNAIPILS++KIEVT    + LTGSNGQISIENTIP
Sbjct: 1   MIQFSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSNGQISIENTIPV 60

Query: 61  SNENAGLLVTNPGSILLEAGFFINIISSLPDVTLEFTEIEQHQIVLTSGKSEITLKGKDV 120
            SNENAGLL+T+PG+ILLEA FFINIISSLPD+++   EIEQHQ+VLTSGKSEITLKGKDV
Sbjct: 61  SNENAGLLITSPGAILLEASFFINIISSLPDISINVKEIEQHQVVLTSGKSEITLKGKDV 120

Query: 121 DQYPRLQEMTTDTPLTLETKLLKSIINETAFAASQQESRPILTGVHLVISQNKYFKAVAT 180
            DQYPRLQE++T+ PL L+TKLLKSII ETAFAAS QESRPILTGVH+V+S +K FKAVAT
Sbjct: 121 DQYPRLQEVSTENPLILKTKLLKSIIAETAFAASLQESRPILTGVHIVLSNHKDFKAVAT 180

Query: 181 DSHRMSQRTFQLEKSANNFDLVVPSKSLREFSAVFTDDIETVEVFFSDSQMLFRSENISF 240
            DSHRMSQR   L+ ++ +FD+V+PSKSLREFSAVFTDDIETVEVFFS SQ+LFRSE+ISF
Sbjct: 181 DSHRMSQRLITLDNTSADFDVVIPSKSLREFSAVFTDDIETVEVFFSPSQILFRSEHISF 240

Query: 241 YTRLLEGNYPDTDRLLTNQFETEIIFNTNALRHAMERAYLISNATQNGTVRLEIQNETVS 300
            YTRLLEGNYPDTDRLL  +FETE++FNT +LRHAMERA+LISNATQNGTV+LEI     +S
Sbjct: 241 YTRLLEGNYPDTDRLLMTEFETEVVFNTQSLRHAMERAFLISNATQNGTVKLEITQNHIS 300
```

```
Query: 301 AHVNSPEVGKVNEELDTVSLKGDSLNISFNPTYLIESLKAVKSETVTIRFISPVRPFTLT 360
           AHVNSPEVGKVNE+LD VS  G  L ISFNPTYLIESLKA+KSETV I F+SPVRPFTLT
Sbjct: 301 AHVNSPEVGKVNEDLDIVSQSGSDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360

Query: 361 PGEDTEDFIQLITPVRTN 378
           PG++ E FIQLITPVRTN
Sbjct: 361 PGDEEESFIQLITPVRTN 378
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1652

A DNA sequence (GBSx1747) was identified in *S.agalactiae* <SEQ ID 5117> which encodes the amino acid sequence <SEQ ID 5118>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0857(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10299> which encodes amino acid sequence <SEQ ID 10300> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00282 GB:AF008220 YtlR [Bacillus subtilis]
 Identities = 83/298 (27%), Positives = 138/298 (45%), Gaps = 35/298 (11%)

Query: 19  YIIANPHAGNKNASTIVGKIQE--LYHTEDISVFYTEQKDDEK--KQVINILRSFKESDH 74
           + I NP AG++N   +   IQ+  +    +   F TE     +   + I+ ++ +K
Sbjct: 5   FFIINPTAGHRNGLRVWKSIQKELIKRKVEHRSFLTEHPGHAEVLARQISTIQEYKLK-R 63

Query: 75  LMIIGGDGTLSKVMTYLPQ--HIPCTYYPVGSGNDFARALKIPNL---------KETLTA 123
           L++IGGDGT+ +V+  L     I  ++ P G+ NDF+R   I  +         K  LT
Sbjct: 64  LIVIGGDGTMHEVVNGLKDVDDIELSFVPAGAYNDFSRGFSIKKIDLIQEIKKVKRPLT- 122

Query: 124 IQTERLKEINCFIYDKGLIL---NSLDLGFAAYVVWKASNSKIKNILNRYRLGKITYIVI 180
            +T  L  +N F+ DK  IL   N + +GF AYV  KA    ++ +    RL  + Y +
Sbjct: 123 -RTFHLGSVN-FLQDKSQILYFMNHIGIGFDAYVNKKAMEFPLRRVFLFLRLRFLVYPL- 179

Query: 181 AIKSLLHSSK------VQVLVEGETGQQIKLNDLYFFALANNTYFGGGITIWPKASALTA 234
              S LH+S       +    E ET +    +D++F  ++N+ ++GGG+   P A+
Sbjct: 180 ---SHLHASATFKPFTLACTTEDETRE---FHDVWFAVVSNHPFYGGGMKAAPLANPREK 233

Query: 235 ELDMVYAKGHTFLKRLSILLSLVFKRHTTSKSIKHQTFKAMTVYFPKNSLIEIDGEIV 292
             D+V  +   FLK+  +L  + F +HT    +     K +T Y         DGEI+
Sbjct: 234 TFDIVIVENQPFLKKYWLLCLMAFGKHTKMDGVTMFKAKDITFYTKDKIPFHADGEIM 291
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1653

A DNA sequence (GBSx1748) was identified in *S.agalactiae* <SEQ ID 5121> which encodes the amino acid sequence <SEQ ID 5122>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3792(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45338 GB:AF000658 ORFX [Streptococcus pneumoniae]
 Identities = 46/63 (73%), Positives = 57/63 (90%)

Query: 1  MYQVGSLVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
           MYQVG+ VEMKKPHAC IK TGKKAN+W++ RVGADIKI+C+NC+HV+MM RYDFERK+
Sbjct: 1  MYQVGNFVEMKKPHACTIKSTGKKANRWEITRVGADIKIKCSNCEHVVMMGRYDFERKMN 60

Query: 61 KVL 63
           K++
Sbjct: 61 KII 63
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5123> which encodes the amino acid sequence <SEQ ID 5124>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4038(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 63/65 (96%), Positives = 64/65 (97%)

Query: 1  MYQVGSLVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
           MYQ+GS VEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK
Sbjct: 1  MYQIGSFVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60

Query: 61 KVLQP 65
           KVLQP
Sbjct: 61 KVLQP 65
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1654

A DNA sequence (GBSx1749) was identified in *S.agalactiae* <SEQ ID 5125> which encodes the amino acid sequence <SEQ ID 5126>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.99    Transmembrane   48 -  64 (  47 -  66)

----- Final Results -----
               bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1655

A DNA sequence (GBSx1750) was identified in *S.agalactiae* <SEQ ID 5127> which encodes the amino acid sequence <SEQ ID 5128>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4171(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1656

A DNA sequence (GBSx1751) was identified in *S.agalactiae* <SEQ ID 5129> which encodes the amino acid sequence <SEQ ID 5130>. This protein is predicted to be GTP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3952(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8875> which encodes amino acid sequence <SEQ ID 8876> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      0.53
GvH: Signal Score (-7.5): -0.13
     Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   1.48 threshold:  0.0
   PERIPHERAL  Likelihood =  1.48     195
 modified ALOM score:  -0.80

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07770 GB:AP001520 GTP-binding protein [Bacillus halodurans]
 Identities = 223/329 (67%), Positives = 273/329 (82%), Gaps = 5/329 (1%)

Query: 1    MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGASKGEGLGNKFLANIREVDAIVH 60
```

```
           +VEVPD RLQKLTEL+ PKKTVPT FEFTDIAGIV+GASKGEGLGN+FL++IR+VDAI H
Sbjct: 43  IVEVPDPRLQKLTELVNPKKTVPTAFEFTDIAGIVEGASKGEGLGNQFLSHIRQVDAISH 102

Query: 61  VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE 120
           VVR FDDEN+    G     VDPI DI  INLELILADLES++KR++RV+K+A+T KDKE
Sbjct: 103 VVRCFDDENITHVSGS----VDPIRDISVINLELILADLESVDKRFSRVQKLAKT-KDKE 157

Query: 121 SVAEFNVLQKIKPVLEDGKSARTIEFTEEEAKVVKGLFLLTTKPVLYVANVDEDKVADPD 180
           +VAE  VL+K+K   E+ K AR+IEFTEE+ K+VKGL LLT+KPVLYVANV ED V  PD
Sbjct: 158 AVAELEVLEKLKDAFENEKPARSIEFTEEQQKIVKGLHLLTSKPVLYVANVSEDDVLSPD 217

Query: 181 DIDYVNQIRAFAETENAEVVVISARAEEEISELDDEDKLEFLEAIGLTESGVDKLTRAAY 240
           D  +V +++AFA  EN+EV+V+SA+ EEEI+ELD E+K  FLE +G+ ESG+D+L RAAY
Sbjct: 218 DNPFVQKVKAFAAEENSEVIVVSAKIEEEIAELDGEEKAMFLEELGIQESGLDQLIRAAY 277

Query: 241 HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAASIIHSDFERGFIRAVTMSYDDLIQYGSEK 300
            LLGL TYFTAGE+EVRAWTF++G KAPQAA IIHSDFE+GFIRA T+SY+DL++ GS
Sbjct: 278 SLLGLQTYFTAGEQEVRAWTFRKGTKAPQAAGIIHSDFEKGFIRAETVSYNDLVEAGSMA 337

Query: 301 AVKEAGRLREEGKEYIVQDGDIMEFRFNV 329
             KE G++R EGKEY+VQDGD++ FRFNV
Sbjct: 338 VAKERGKVRLEGKEYVVQDGDVIHFRFNV 366
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5131> which encodes the amino acid sequence <SEQ ID 5132>. Analysis of this protein sequence reveals the following:

```
        Possible site: 29

  >>> Seems to have a cleavable N-term signal seq.

  ----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB07770 GB:AP001520 GTP-binding protein [Bacillus halodurans]
 Identities = 259/371 (69%), Positives = 314/371 (83%), Gaps = 5/371 (1%)

Query: 1   MALTAGIVGLPNVGKSTLFNAITKAGAEAANYPFATIDPNVGMVEVPDERLQKLTELITP 60
           MALT GIVGLPNVGKSTLFNAIT+AGAE+ANYPF TIDPNVG+VEVPD RLQKLTEL+ P
Sbjct: 1   MALTTGIVGLPNVGKSTLFNAITQAGAESANYPFCTIDPNVGIVEVPDPRLQKLTELVNP 60

Query: 61  KKTVPTTFEFTDIAGIVKGASRGEGLGNKFLANIREIDAIVHVVRAFDDENVMREQGRED 120
           KKTVPT FEFTDIAGIV+GAS+GEGLGN+FL++IR++DAI HVVR FDDEN+    G
Sbjct: 61  KKTVPTAFEFTDIAGIVEGASKGEGLGNQFLSHIRQVDAISHVVRCFDDENITHVSGS-- 118

Query: 121 AFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKESVAEFNVLQKIKPVLEDG 180
             VDPI DI  INLELILADLES++KR++RV+K+A+T KDKE+VAE  VL+K+K   E+
Sbjct: 119 --VDPIRDISVINLELILADLESVDKRFSRVQKLAKT-KDKEAVAELEVLEKLKDAFENE 175

Query: 181 KSARTIEFTEDEAKVVKGLFLLTTKPVLYVANVDEDKVANPDGIDYVKQIRDFAATENAE 240
           K AR+IEFTE++ K+VKGL LLT+KPVLYVANV ED V +PD   +V++++ FAA EN+E
Sbjct: 176 KPARSIEFTEEQQKIVKGLHLLTSKPVLYVANVSEDDVLSPDDNPFVQKVKAFAAEENSE 235

Query: 241 VVVISARAEEEISELDDEDKEEFLEAIGLTESGVDKLTRAAYHLLGLGTYFTAGEKEVRA 300
           V+V+SA+ EEEI+ELD E+K  FLE +G+ ESG+D+L RAAY LLGL TYFTAGE+EVRA
Sbjct: 236 VIVVSAKIEEEIAELDGEEKAMFLEELGIQESGLDQLIRAAYSLLGLQTYFTAGEQEVRA 295

Query: 301 WTFKRGIKAPQAAGIIHSDFERGFIRAVTMSYDDLMTYGSEKAVKEAGRLREEGKEYVVQ 360
           WTF++G KAPQAAGIIHSDFE+GFIRA T+SY+DL+  GS    KE G++R EGKEYVVQ
Sbjct: 296 WTFRKGTKAPQAAGIIHSDFEKGFIRAETVSYNDLVEAGSMAVAKERGKVRLEGKEYVVQ 355

Query: 361 DGDIMEFRFNV 371
           DGD++ FRFNV
Sbjct: 356 DGDVIHFRFNV 366
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 316/329 (96%), Positives = 322/329 (97%)

Query: 1   MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGASKGEGLGNKFLANIREVDAIVH 60
            MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGAS+GEGLGNKFLANIRE+DAIVH
Sbjct: 43  MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGASRGEGLGNKFLANIREIDAIVH 102

Query: 61  VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE 120
            VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE
Sbjct: 103 VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE 162

Query: 121 SVAEFNVLQKIKPVLEDGKSARTIEFTEEEAKVVKGLFLLTTKPVLYVANVDEDKVADPD 180
            SVAEFNVLQKIKPVLEDGKSARTIEFTE+EAKVVKGLFLLTTKPVLYVANVDEDKVA+PD
Sbjct: 163 SVAEFNVLQKIKPVLEDGKSARTIEFTEDEAKVVKGLFLLTTKPVLYVANVDEDKVANPD 222

Query: 181 DIDYVNQIRAFAETENAEVVVISARAEEEISELDDEDKLEFLEAIGLTESGVDKLTRAAY 240
             IDYV QIR FA TENAEVVVISARAEEEISELDDEDK EFLEAIGLTESGVDKLTRAAY
Sbjct: 223 GIDYVKQIRDFAATENAEVVVISARAEEEISELDDEDKEEFLEAIGLTESGVDKLTRAAY 282

Query: 241 HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAASIIHSDFERGFIRAVTMSYDDLIQYGSEK 300
            HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAA IIHSDFERGFIRAVTMSYDDL+ YGSEK
Sbjct: 283 HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAAGIIHSDFERGFIRAVTMSYDDLMTYGSEK 342

Query: 301 AVKEAGRLREEGKEYIVQDGDIMEFRFNV 329
            AVKEAGRLREEGKEY+VQDGDIMEFRFNV
Sbjct: 343 AVKEAGRLREEGKEYVVQDGDIMEFRFNV 371
```

SEQ ID 8876 (GBS177) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 4; MW 41.2kDa).

The GBS177-His fusion product was purified (Figure 118A; see also Figure 202, lane 7) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1657

A DNA sequence (GBSx1752) was identified in *S.agalactiae* <SEQ ID 5133> which encodes the amino acid sequence <SEQ ID 5134>. This protein is predicted to be stage V sporulation protein C (pth). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2212(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10301> which encodes amino acid sequence <SEQ ID 10302> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03787 GB:AP001507 stage V sporulation protein C
          (peptidyl-tRNA hydrolase) [Bacillus halodurans]
```

```
 Identities = 89/187 (47%), Positives = 127/187 (67%), Gaps = 2/187 (1%)

Query: 6   VKMIVGLGNPGSKYNDTKHNIGFMAVDRIVKDLDVNFTEDKNFKAEIGSDFINGEKIYFI 65
           +K+IVGLGNPG+KY+ T+HN+GF  VD + + L++   + K      G   I+GEKI+ +
Sbjct: 1   MKLIVGLGNPGAKYDGTRHNVGFDVVDAVARRLNIEIKQSKA-NGLYGEGRIDGEKIFLL 59

Query: 66  KPTTFMNNSGIAVKALLTYYNISIKDMIIIYDDLDMEVGKIRFRQKGSAGGHNGIKSIIA 125
           KP TFMN SG +V+  L YYN+ ++D+++IYDDLD+ VGKIR RQKGSAGGHNG+KS+IA
Sbjct: 60  KPQTFMNRSGESVRPFLEYYNMEVEDLLVIYDDLDLPVGKIRLRQKGSAGGHNGMKSLIA 119

Query: 126 HLGTQEFDRIKVGIGRPNGRMTVINHVLGKFDKNDEIMILNTLDKVDNAVNYYLQTNDFQ 185
           HLGT +F RI+VG+ RP    TV+ HVLG++   ++  I   +D    A   + +   F
Sbjct: 120 HLGTSDFKRIRVGVDRPAPGETVVQHVLGRYRPEEKDAISEAIDLSAEAAEAFTK-KPFL 178

Query: 186 KTMQKYN 192
           + M  +N
Sbjct: 179 EVMNTFN 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5135> which encodes the amino acid sequence <SEQ ID 5136>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 148/189 (78%), Positives = 166/189 (87%)

Query: 5   MVKMIVGLGNPGSKYNDTKHNIGFMAVDRIVKDLDVNFTEDKNFKAEIGSDFINGEKIYF 64
           MVKMIVGLGNPGSKY  TKHNIGFMA+D IVK+LDV FT+DKNFKA+IGS FIN EK+YF
Sbjct: 16  MVKMIVGLGNPGSKYEKTKHNIGFMAIDNIVKNLDVTFTDDKNFKAQIGSTFINHEKVYF 75

Query: 65  IKPTTFMNNSGIAVKALLTYYNISIKDMIIIYDDLDMEVGKIRFRQKGSAGGHNGIKSII 124
           +KPTTFMNNSGIAVKALLTYYNI I D+I+IYDDLDMEV K+R R KGSAGGHNGIKSII
Sbjct: 76  VKPTTFMNNSGIAVKALLTYYNIDITDLIVIYDDLDMEVSKLRLRSKGSAGGHNGIKSII 135

Query: 125 AHLGTQEFDRIKVGIGRPNGRMTVINHVLGKFDKNDEIMILNTLDKVDNAVNYYLQTNDF 184
           AH+GTQEF+RIKVGIGRP   MTVINHV+G+F+  D I I  TLD+V NAV +YLQ NDF
Sbjct: 136 AHIGTQEFNRIKVGIGRPLKGMTVINHVMGQFNTEDNIAISLTLDRVVNAVKFYLQENDF 195

Query: 185 QKTMQKYNG 193
           +KTMQK+NG
Sbjct: 196 EKTMQKFNG 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1658

A DNA sequence (GBSx1753) was identified in *S.agalactiae* <SEQ ID 5137> which encodes the amino acid sequence <SEQ ID 5138>. This protein is predicted to be transcription-repair coupling factor (mfd). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2456(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD03810 GB:AF054624 transcription-repair coupling factor
          [Lactobacillus sakei]
 Identities = 523/1051 (49%), Positives = 733/1051 (68%), Gaps = 20/1051 (1%)

Query: 1    MNIIELFSQNKVVRTWHSGLVTNSRQLVMGFSGASKAIAIASAYEKLSKKIMVVTATQTD 60
            M++I +    + V++         RQL+ G SG++K + +A+ Y++  + ++++ +
Sbjct: 1    MDLISMLGNTQQVQSVLENQKPGVRQLLTGLSGSAKTLFLATIYKQQRQPLLIIESNMFQ 60

Query: 61   SDKLSSDISSLIGEDNVYQFFADDVPAAEFIFSSLDKSISRLSALRFLKDPEKNGVLITS 120
            +++++ D+++ +  D +Y F  ++V AAE   SS +    R+  L FL   +K G+++TS
Sbjct: 61   ANQVAEDLANQLNGDQIYTFPVEEVMAAEIAVSSPESRAERVRTLSFLATGKK-GIVVTS 119

Query: 121  ISGLRLLLPNPEVFSKSQYKFEIGQECYLDKLCKNLVNLGYQKVSQVFSPGEFSQRGDIL 180
            ++G+R LLP    +  SQ + E+G E     L   L  +GY +   V  PGEF+ RGDI+
Sbjct: 120  VAGMRRLLPTVRQWRDSQTQIEMGGEVDPKILGAQLAEMGYHRDKLVGKPGEFAMRGDII 179

Query: 181  DIFEMTQEYPYRLEFFGDEIDGIRQFDIDTQKSLKQLESVQISPADDIILQDADFERAKK 240
            DIF +  E P R+E F  E+D IR F+ DTQ+S++ LESV I PA D++   A  E A +
Sbjct: 180  DIFPLDTENPVRIELFDTEVDAIRSFEADTQRSIENLESVAIMPATDLLANAAQLEMAGE 239

Query: 241  KLEG-YLVTASEVQ------------RTYLSEVLSTTENHFKHSDIRRFLSIFYEKEWGI 287
             L+  Y  TA+++              T +S +L+      +  ++  F+   Y     +
Sbjct: 240  ALQADYQQTAAKITAKDDQKALAVNFETPISRLLAGE----RLENLALFVDYLYPDHTSL 295

Query: 288  LDYIPEGTPLFVDDFQKIVDRNAKLDLEIASLLTEDLQQGKSHSSLNYFSDPYKQLRQYQ 347
            +DY      +  DD+ +I +    L  E A+  T+ L   +   +     D +  ++Q Q
Sbjct: 296  IDYFKNSGLVVFDDYPRIQETQRVLAEEAANWQTDMLGSRRLLPAQKLLVDVHHLMKQDQ 355

Query: 348  -PATFFSNFHKGLGNLKFDKLHHFTQYGMQEFFNQFPLLVDEINRYKKSGATVLLQVDSQ 406
             P  + S F KG+G LK D L +      +Q+FF+Q PLL  E++R++K   TV++ V
Sbjct: 356  HPHLYLSLFQKGMGKLKLDTLGNMPTRNVQQFFSQMPLLKTEMSRWQKQQQTVVVLVSDA 415

Query: 407  KGLNLLQENLKEYGLDLIISDKNDIVQKESQLIVGHLSNGFYFADEKIVLITEREIYHRR 466
            K +  + +   ++ ++  ++ K  +V  + Q++ G L NGF   D K+V++TE+E+++
Sbjct: 416  KRVKKIDQTFHDFEIEATVTTKTKLVAGQIQIVQGSLQNGFELPDLKLVVLTEKELFNTA 475

Query: 467  VKRKIRRSNISNAERLKDYNELSVGDYVVHNVHGVGKFLGIETIEIQGIHRDYLTIQYQN 526
             K+K+RR  ++NAERLK Y+EL  GDYVVH  HG+G+++G+ET+E+ G+H+DY+TI Y++
Sbjct: 476  PKKKVRRQTLANAERLKSYSELKPGDYVVHVNHGIGEYVGMETLEVDGVHQDYITILYRD 535

Query: 527  ADRISIPVEQIELLTKYVSADGKEPKINTLNDGRFKKAKQRVAKQVEDIADDLLKLYAER 586
              ++ IPV Q++++ KYVSA+ K PKIN L    ++K K +V+ ++EDIADDL++LYA+R
Sbjct: 536  NGKLFIPVTQLDMVQKYVSAESKTPKINKLGGAEWQKTKSKVSAKIEDIADDLIELYAQR 595

Query: 587  SQLQGFAFSPDDNMQNDFDNDFAYVETEDQLRSIKEIKQDMEGNRPMDRLLVGDVGFGKT 646
               +G+AF  DD +Q DF+N FAY ET+DQLRS  EIK DME  RPMDRLLVGDVGFGKT
Sbjct: 596  EAEKGYAFPKDDQLQADFENQFAYPETDDQLRSTAEIKHDMEKVRPMDRLLVGDVGFGKT 655

Query: 647  EVAMRAAFKAVNDHKQVVVLVPTTVLAQQHFENFKERFSNYPVTVDVLSRFRSKKEQTDT 706
            EVA+RAAFKAV   KQV  LVPTT+LAQQH+EN   RF+++PV + +LSRF+++KE T T
Sbjct: 656  EVALRAAFKAVAAGKQVAFLVPTTILAQQHYENMLARFADFPVELGLLSRFKTRKEVTAT 715

Query: 707  LKRLSKGQVDIIIGTHRLLSQDVVFSDLGLIVIDEEQRFGVKHKEKLKELKTKVDVLTLT 766
            LK L KGQVDI+IGTHRLLS+DVVF DLGL+++DEEQRFGVKHKE+LK+LK +VDVLTLT
Sbjct: 716  LKGLEKGQVDIVIGTHRLLSKDVVFKDLGLLIVDEEQRFGVKHKERLKQLKAQVDVLTLT 775

Query: 767  ATPIPRTLHMSMLGIRDLSVIETPPTNRYPVQTYVLETNPGLVREAIIREIDRGGQVFYV 826
            ATPIPRTLHMSMLG+RDLSVIETPPTNRYP+QTYV+E N G +REAI RE++R GQVFY+
Sbjct: 776  ATPIPRTLHMSMLGVRDLSVIETPPTNRYPIQTYVMEQNAGAMREAIERELERNGQVFYL 835

Query: 827  YNKVDTIDQKVSELQELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIETGV 886
            +N+V  I+Q V E+Q LVPEA++G+ HGQM+E QLE  + DF+ G YDVLV TTIIETGV
Sbjct: 836  HNRVSDIEQTVDEIQALVPEATVGYAHGQMTEAQLEGVIYDFVQGKYDVLVTTTIIETGV 895
```

```
Query: 887  DISNVNTLFVENADHMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEISEKRLDAIKG 946
             D+ NVNT+ VE+ADH GLS LYQLRGR+GRS+R+AY Y MY+PDKVLTE+SEKRL AIK
Sbjct: 896  DMPNVNTMIVEDADHYGLSQLYQLRGRIGRSSRVAYGYFMYKPDKVLTEVSEKRLQAIKD 955

Query: 947  FTELGSGFKIAMRDLSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIATKQGKSLIRQK 1006
             FTELGSGFKIAMRDLSIRGAGN+LG  Q GFIDSVGF++YSQ+L +A+A KQGK  +  K
Sbjct: 956  FTELGSGFKIAMRDLSIRGAGNLLGKQQHGFIDSVGFDLYSQMLSEAVAKKQGKK-VAAK 1014

Query: 1007 GNAELALQIDAYLPAEYISDERQKIEIYKRI 1037
              NAE+ L+++AYLP +YI+D+RQKIEIYKRI
Sbjct: 1015 TNAEIDLKLEAYLPDDYINDQRQKIEIYKRI 1045
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5139> which encodes the amino acid sequence <SEQ ID 5140>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2826(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 875/1161 (75%), Positives = 1032/1161 (88%)

Query: 1    MNIIELFSQNKVVRTWHSGLVTNSRQLVMGFSGASKAIAIASAYEKLSKKIMVVTATQTD 60
            M+I+ELFSQNK V++WHSGL T  RQLVMG SG+SK +AIASAY    KKI+VVT+TQ +
Sbjct: 1    MDILELFSQNKKVQSWHSGLTTLGRQLVMGLSGSSKTLAIASAYLDDQKKIVVVTSTQNE 60

Query: 61   SDKLSSDISSLIGEDNVYQFFADDVPAAEFIFSSLDKSISRLSALRFLKDPEKNGVLITS 120
             +KL+SD+SSL+ E+ V+QFFADDV AAEFIF+S+DK++SR+  L+FL++P+  GVLI S
Sbjct: 61   VEKLASDLSSLLDEELVFQFFADDVAAAEFIFASMDKALSRIETLQFLRNPKSQGVLIVS 120

Query: 121  ISGLRLLLPNPEVFSKSQYKFEIGQECYLDKLCKNLVNLGYQKVSQVFSPGEFSQRGDIL 180
            +SGLR+LLPNP+VF+KSQ +  +G++   D L K L+ +GYQKVSQV SPGEFS+RGDIL
Sbjct: 121  LSGLRILLPNPDVFTKSQIQLTVGEDYDSDTLTKQLMTIGYQKVSQVISPGEFSRRGDIL 180

Query: 181  DIFEMTQEYPYRLEFFGDEIDGIRQFDIDTQKSLKQLESVQISPADDIILQDADFERAKK 240
            DI+E+TQE PYRLEFFGD+ID IRQF  +TQKS +QLE + I+PA D+I + +DF+R  +
Sbjct: 181  DIYEITQELPYRLEFFGDDIDSIRQFHPETQKSFEQLEGIFINPASDLIFEVSDFQRGIE 240

Query: 241  KLEGYLVTASEVQRTYLSEVLSTTENHFKHSDIRRFLSIFYEKEWGILDYIPEGTPLFVD 300
            +LE  L TA + +++YL +VL+ ++N FKH DIR+F S+FYEKEW +LDYIP+GTP+F D
Sbjct: 241  QLEKALQTAQDDKKSYLEDVLAVSKNGFKHKDIRKFQSLFYEKEWSLLDYIPKGTPIFFD 300

Query: 301  DFQKIVDRNAKLDLEIASLLTEDLQQGKSHSSLNYFSDPYKQLRQYQPATFFSNFHKGLG 360
            DFQK+VD+NA+ DLEIA+LLTEDLQQGK+ S+LNYF+D Y++LR Y+PATFFSNFHKGLG
Sbjct: 301  DFQKLVDKNARFDLEIANLLTEDLQQGKALSNLNYFTDNYRELRHYKPATFFSNFHKGLG 360

Query: 361  NLKFDKLHHFTQYGMQEFFNQFPLLVDEINRYKKSGATVLLQVDSQKGLNLLQENLKEYG 420
            N+KFD++H  TQY MQEFFNQFPLL+DEI RY+K+  TV++QV+SQ     L+++ ++Y
Sbjct: 361  NIKFDQMHQLTQYAMQEFFNQFPLLIDEIKRYQKNQTTVIVQVESQYAYERLEKSFQDYQ 420

Query: 421  LDLIISDKNDIVQKESQLIVGHLSNGFYFADEKIVLITEREIYHRRVKRKIRRSNISNAE 480
              L +   N IV +ESQ+++G +S+GFYFADEK+ LITE EIYH+++KR+ RRSNISNAE
Sbjct: 421  FRLPLVSANQIVSRESQIVIGAISSGFYFADEKLALITEHEIYHKKIKRRARRSNISNAE 480

Query: 481  RLKDYNELSVGDYVVHNVHGVGKFLGIETIEIQGIHRDYLTIQYQNADRISIPVEQIELL 540
            RLKDYNEL+VGDYVVHNVHG+G+FLGIETI+IQGIHRDY+TIQYQN+DRIS+P++QI  L
Sbjct: 481  RLKDYNELAVGDYVVHNVHGIGRFLGIETIQIQGIHRDYVTIQYQNSDRISLPIDQISSL 540

Query: 541  TKYVSADGKEPKINTLNDGRFKKAKQRVAKQVEDIADDLLKLYAERSQLQGFAFSPDDNM 600
            +KYVSADGKEPKIN LNDGRF+K KQ+VA+QVEDIADDLLKLYAERSQ +GF+FSPDD++
Sbjct: 541  SKYVSADGKEPKINKLNDGRFQKTKQKVARQVEDIADDLLKLYAERSQQKGFSFSPDDDL 600
```

```
Query: 601  QNDFDNDFAYVETEDQLRSIKEIKQDMEGNRPMDRLLVGDVGFGKTEVAMRAAFKAVNDH 660
             Q  FD+DFA+VETEDQLRSIKEIK DME  +PMDRLLVGDVGFGKTEVAMRAAFKAVNDH
Sbjct: 601  QRAFDDDFAFVETEDQLRSIKEIKADMESMQPMDRLLVGDVGFGKTEVAMRAAFKAVNDH 660

Query: 661  KQVVVLVPTTVLAQQHFENFKERFSNYPVTVDVLSRFRSKKEQTDTLKRLSKGQVDIIIG 720
             KQV VLVPTTVLAQQH+ENFK RF NYPV VDVLSRFRSKKEQ +TL+R+ KGQ+DIIIG
Sbjct: 661  KQVAVLVPTTVLAQQHYENFKARFENYPVEVDVLSRFRSKKEQAETLERVRKGQIDIIIG 720

Query: 721  THRLLSQDVVFSDLGLIVIDEEQRFGVKHKEKLKELKTKVDVLTLTATPIPRTLHMSMLG 780
             THRLLS+DVVFSDLGLIVIDEEQRFGVKHKE LKELKTKVDVLTLTATPIPRTLHMSMLG
Sbjct: 721  THRLLSKDVVFSDLGLIVIDEEQRFGVKHKETLKELKTKVDVLTLTATPIPRTLHMSMLG 780

Query: 781  IRDLSVIETPPTNRYPVQTYVLETNPGLVREAIIREIDRGGQVFYVYNKVDTIDQKVSEL 840
             IRDLSVIETPPTNRYPVQTYVLE NPGLVREAIIRE+DRGGQ+FYVYNKVDTI++KV+EL
Sbjct: 781  IRDLSVIETPPTNRYPVQTYVLENNPGLVREAIIREMDRGGQIFYVYNKVDTIEKKVAEL 840

Query: 841  QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIETGVDISNVNTLFVENAD 900
             QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIETGVDISNVNTLF+ENAD
Sbjct: 841  QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIETGVDISNVNTLFIENAD 900

Query: 901  HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEISEKRLDAIKGFTELGSGFKIAMRD 960
             HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTE+SEKRL+AIKGFTELGSGFKIAMRD
Sbjct: 901  HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEVSEKRLEAIKGFTELGSGFKIAMRD 960

Query: 961  LSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIATKQGKSLIRQKGNAELALQIDAYLP 1020
             LSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIA+KQGK+ +RQKGN E+ LQIDAYLP
Sbjct: 961  LSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIASKQGKTTVRQKGNTEINLQIDAYLP 1020

Query: 1021 AEYISDERQKIEIYKRIRELETRADYEALQDELIDRFGEYPDQVAYLLEIGLLKAYLDLA 1080
              +YI+DERQKI+IYKRIRE+++R DY  LQDEL+DRFGEYPDQVAYLLEI LLK Y+D A
Sbjct: 1021 DDYIADERQKIDIYKRIREIQSREDYLNLQDELMDRFGEYPDQVAYLLEIALLKHYMDNA 1080

Query: 1081 FTELVERKGNEISILFEKASLKYFLTQDYFEALSKTQLKARISETNGKMEVVFNIKHKKN 1140
             F ELVERK N++ + FE  SL YFLTQDYFEALSKT LKA+ISE  GK+++VF+++H+K+
Sbjct: 1081 FAELVERKNNQVIVRFEVTSLTYFLTQDYFEALSKTHLKAKISEHQGKIDIVFDVRHQKD 1140

Query: 1141 YEIIEELLKFAECFIEIKSRK 1161
             Y I+EEL+ F E   EIK RK
Sbjct: 1141 YRILEELMLFGERLSEIKIRK 1161
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1659

A DNA sequence (GBSx1754) was identified in *S.agalactiae* <SEQ ID 5141> which encodes the amino acid sequence <SEQ ID 5142>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4347(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11835 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 50/84 (59%), Positives = 70/84 (82%)

Query: 1  MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDLKLNDQVEIRFGNKLLTVKV 60
           MRLDK+LKVSR+IKRR +AKEVAD+GR+ +NG  AK+S+D+K  D++ +RFG KL+TV+V
Sbjct: 1  MRLDKFLKVSRLIKRRTLAKEVADQGRISINGNQAKASSDVKPGDELTVRFGQKLVTVQV 60

Query: 61 LEMKDSTKKEDAIKMYEIINETRI 84
```

```
           E+KD+TKKE+A  MY I+ E ++
Sbjct: 61 NELKDTTKKEEAANMYTILKEEKL 84
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5143> which encodes the amino acid sequence <SEQ ID 5144>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2963(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 72/90 (80%), Positives = 85/90 (94%)

Query: 1  MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDLKLNDQVEIRFGNKLLTVKV 60
           MRLDKYLKVSR+IKRR VAKEVADKGR+KVNG+LAKSST++KLND +EI FGNKLLTV+V
Sbjct: 9  MRLDKYLKVSRLIKRRSVAKEVADKGRIKVNGILAKSSTNIKLNDHIEISFGNKLLTVRV 68

Query: 61 LEMKDSTKKEDAIKMYEIINETRIETDEQA 90
           +E+KDSTKKEDA+KMYEII+ETRI  +E+A
Sbjct: 69 IEIKDSTKKEDALKMYEIISETRITLNEEA 98
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1660

A DNA sequence (GBSx1755) was identified in *S.agalactiae* <SEQ ID 5145> which encodes the amino acid sequence <SEQ ID 5146>. This protein is predicted to be DivIC homolog. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12   Transmembrane   34 -  50 (  31 -  56)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98903 GB:AF023181 DivIC homolog [Listeria monocytogenes]
 Identities = 36/119 (30%), Positives = 65/119 (54%), Gaps = 2/119 (1%)

Query: 2   SKPNVVQLNNQYINDE-NLKKRYEAEELRRKNRLMGWVLIFVMLLFILPTYNLVKSYRTL 60
            +K  V ++ N+YI D   +KK      +    RL    +IF ++  +L T    K   TL
Sbjct: 4   AKSKVARIENRYIKDTATMKKTRSRRRIALFRRLAFMAIIFAVVGGLL-TITYTKQVLTL 62

Query: 61  QERRQEVVKLTKDYQTLTNRTENQKLLAKQLKNPDYVQKYARAKYYFSKTGEMIYPLPD 119
            +E++++ V++ K    + +  ++      K+L N DY+ K AR++YY SK GE+I+ +P+
Sbjct: 63  KEKKEKQVQVDKKMVAMKDEQDSLNEQIKKLHNDDYIAKLARSEYYLSKDGEIIFNIPE 121
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5147> which encodes the amino acid sequence <SEQ ID 5148>. Analysis of this protein sequence reveals the following:

```
     Possible site: 50

>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -3.93   Transmembrane   34 -  50 (  32 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC98903 GB:AF023181 DivIC homolog [Listeria monocytogenes]
 Identities = 27/116 (23%), Positives = 59/116 (50%)

Query: 3   KPSIVQLNNHYIKKENLKKKFEEEESQKRNRFMGWILVSMMFLFILPTYNLVKSYVDFEK 62
            K  + ++ N YIK     KK          R + ++ +    +  L T    K  +  ++
Sbjct: 5   KSKVARIENRYIKDTATMKKTRSRRRIALFRRLAFMAIIFAVVGGLLTITYTKQVLTLKE 64

Query: 63  QNQQVVKLKKEYNELSESTKKEKQLAERLKDDNFVKKYARAKYYLSREGEMIYPIP 118
            + ++ V++ K+   + +      +  ++L +D+++ K AR++YYLS++GE+I+ IP
Sbjct: 65  KKEKQVQVDKKMVAMKDEQDSLNEQIKKLHNDDYIAKLARSEYYLSKDGEIIFNIP 120
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 73/123 (59%), Positives = 99/123 (80%)

Query: 1   MSKPNVVQLNNQYINDENLKKRYEAEELRRKNRLMGWVLIFVMLLFILPTYNLVKSYRTL 60
            M KP++VQLNN YI  ENLKK++E EE +++NR MGW+L+ +M LFILPTYNLVKSY
Sbjct: 1   MKKPSIVQLNNHYIKKENLKKKFEEEESQKRNRFMGWILVSMMFLFILPTYNLVKSYVDF 60

Query: 61  QERRQEVVKLTKDYQTLTNRTENQKLLAKQLKNPDYVQKYARAKYYFSKTGEMIYPLPDL 120
            +++ Q+VVKL K+Y  L+  T+ +K LA++LK+ ++V+KYARAKYY S+ GEMIYP+P L
Sbjct: 61  EKQNQQVVKLKKEYNELSESTKKEKQLAERLKDDNFVKKYARAKYYLSREGEMIYPIPGL 120

Query: 121 LPK 123
            LPK
Sbjct: 121 LPK 123
```

SEQ ID 5146 (GBS418) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 3; MW 42kDa).

GBS418-GST was purified as shown in Figure 219, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1661

A DNA sequence (GBSx1756) was identified in *S.agalactiae* <SEQ ID 5149> which encodes the amino acid sequence <SEQ ID 5150>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4355(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1662

A DNA sequence (GBSx1757) was identified in *S.agalactiae* <SEQ ID 5151> which encodes the amino acid sequence <SEQ ID 5152>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.52   Transmembrane    4 -  20 (   3 -  22)

----- Final Results -----
               bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5153> which encodes the amino acid sequence <SEQ ID 5154>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 205/428 (47%), Positives = 285/428 (65%)

Query: 1   MKKVLTFLLCSLYFVSIPAISTEEPLTLSQNRRYALTQTVVDKEMYFDAIPERPTTKIEI 60
           M+K+L  +L + +   +P ISTE+ L  S+N  Y L Q VV    +++ IP  P    E
Sbjct: 1   MRKLLAAMLMTFFLTPLPVISTEKKLIFSKNAVYQLKQDVVQSTQFYNQIPSNPNLYQET 60

Query: 61  SSFQDEALTITGETLVPNTLLSIVSLTINSNGIPVFTLSNGQFIKASREAIFNDLVSKQQ 120
            +++D  LT+    L  N  L I SL +N   +PVF L++G +++A+R+ I++D+V  Q
Sbjct: 61  CAYKDSDLTLPAGRLGVNQPLLIKSLVLNKESLPVFELADGTYVEANRQLIYDDIVLNQV 120

Query: 121 SVSLDYWLKPSFVTYEAPYTNGVSEVKNNLKPYSRVHLVEQAETEHGIYYKTDSGFWISV 180
            +   +W +     Y APY  G   + ++     +VH  + A+T HG YY  D   W S
Sbjct: 121 DIDSYFWTQKKLRLYSAPYVLGTQTIPSSFLFAQKVHATQMAQTNHGTYYLIDDKGWASQ 180

Query: 181 EDLSVADNRMAKVQEVLLEKYNKDKYGIYIKQLNTQTVAGINIDRSMYSASIAKLATLYA 240
           EDL   DNRM KVQE+LL+KYN   Y I++KQLNTQT AGIN D+ MY+ASI+KLA LY
Sbjct: 181 EDLVQFDNRMLKVQEMLLQKYNNPNYSIFVKQLNTQTSAGINADKKMYAASISKLAPLYI 240

Query: 241 SQEQVKLGKLSLDSKFEYKDNVNQFPNSYDPSGSGKLEKKADHKLYTVKELLEATAKESD 300
            Q+Q++  KL+ +    Y  +VN F   YDP GSGK+ K AD+K Y V++LL+A A++SD
Sbjct: 241 VQKQLQKKKLAENKTLTYTKDVNHFYGDYDPLGSGKISKIADNKDYRVEDLLKAVAQQSD 300

Query: 301 NVATNMLGYYVNNQYDSMFQTQVDTISGMHWDMKKRQISPQAAGKMMEAIYYQNGDIVNY 360
           NVATN+LGYY+ +QYD  F++++  +SG+ WDM++R ++ ++A  MMEAIY+Q G I++Y
Sbjct: 301 NVATNILGYYLCHQYDKAFRSEIKALSGIDWDMEQRLLTSRSAANMMEAIYHQKGQIISY 360

Query: 361 LSKTDFDNTRIPKNIPVKVAHKIGDAYDYKHDAAIVYAEQPFIMIIFTDKSSYDDITKIA 420
           LS T+FD  RI KNI V VAHKIGDAYDYKHD AIVY   PFI+ IFT+KS+Y+DIT IA
Sbjct: 361 LSNTEFDQQRITKNITVPVAHKIGDAYDYKHDVAIVYGNTPFILSIFTNKSTYEDITAIA 420

Query: 421 DDVYQVLK 428
           DDVY +LK
Sbjct: 421 DDVYGILK 428
```

SEQ ID 5152 (GBS116) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 3; MW 48.5kDa). The GBS116-His fusion product was purified (Figure 202, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 316), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1663

A DNA sequence (GBSx1758) was identified in *S.agalactiae* <SEQ ID 5155> which encodes the amino acid sequence <SEQ ID 5156>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2260(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35664 GB:AE001733 conserved hypothetical protein [Thermotoga maritima]
 Identities = 100/404 (24%), Positives = 181/404 (44%), Gaps = 61/404 (15%)

Query: 19  QKVLIAVSGGIDSINLLQFLYQYQKELSISIGIAHINHGQRKESEKEEEYIRQWGQIHDV 78
           + VL+AVSGGIDS+ LL  L ++   L I I  AH++H  R+ S ++ E++ +  +  ++
Sbjct: 6   EHVLVAVSGGIDSMTLLYVLRKFSPLLKIKITAAHLDHRIRESSRRDREFVERICRQWNI 65

Query: 79  PVFISYF--------QGIFSEDRARNHRYNFFSKVMREEGYTALVTAHHADDQAETVFMR 130
           PV  S           G   E+ AR  RY+F  +  ++ G + +  AHH +D  ETV  R
Sbjct: 66  PVETSEVDVPSLWKDSGKTLEEIAREVRYDFLKRTAKKVGASKIALAHHKNDLLETVVHR 125

Query: 131 ILRGSRLRYLSGIKQVSAFANGQLIRPFLPYKKELLP------NIFHFEDASNASSDYLR 184
           ++RG+    L+ I         + IRPFL +K+  +       N+ +  D +N +  Y R
Sbjct: 126 LIRGTGPLGLACISP----KREEFIRPFLVFKRSEIEEYARKNNVPYVVDETNYNVKYTR 181

Query: 185 NRIRNVYFPALERENNQLKDSLITLSEETECLFTALTDLTRSIEVTNCYDF--------- 235
           N IR+   P ++  N  ++D++  L   T  L   +    +     N Y +
Sbjct: 182 NFIRHRIVPLMKELNPTVEDAVYRLVSVTHLLRNFVERTVQDFVERNVYFYKDYAVFVEP 241

Query: 236 --LRQTHSVQEFLLQDYISKFPDLQVSKEQFRVILKLIRTKANIDYTIKSGYFLHKDYES 293
             L     V  ++L++   + P+ +          KLI T  +    + SG F+ + +
Sbjct: 242 EDLFLFLEVTRWVLKEMYGRVPEYE----------KLIGTLKSKRVELWSGIFVERSFGY 291

Query: 294 FHITKIHPKTDSFKVEKRLELHNIQIFSQYLFSYGKFISQADITIPIYDT---SPIILRR 350
             + K       FK + R+E+             G  +      I + +        +R
Sbjct: 292 VAVGK-----TVFKKKYRVEVK------------GDMLEMEGFKIRVVNNRNDMKFWVRN 334

Query: 351 RKEGDRIFLGNHTKKIRRLFIDEKIT--LKEREEAVIGEQNKEL 392
           RKEGDRI +    +K++ +FI++K+    ++R   ++ E+++ L
Sbjct: 335 RKEGDRIIVNGRERKLKDVFIEKKVPTFYRDRVPLLVDEEDRVL 378
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5157> which encodes the amino acid sequence <SEQ ID 5158>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2187(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 218/424 (51%), Positives = 290/424 (67%), Gaps = 2/424 (0%)

Query: 2   YNTILKDTLSKGLFTAHQKVLIAVSGGIDSINLLQFLYQYQKELSISIGIAHINHGQRKE 61
           Y  I  +  +K  F  H+ VLIAVSGG+DS+NLL FLY +Q +L I IGIAH+NH QR E
Sbjct: 4   YQEIFNEIKNKAYFKNHRHVLIAVSGGVDSMNLLHFLYLFQDKLKIRIGIAHVNHKQRSE 63

Query: 62  SEKEEEYIRQWGQIHDVPVFISYFQGIFSEDRARNHRYNFFSKVMREEGYTALVTAHHAD 121
           S+ EE Y++ W + HD+P+++S F+GIFSE  AR+ RY FF  +M +  Y+ALVTAHH+D
Sbjct: 64  SDSEEAYLKCWAKKHDIPIYVSNFEGIFSEKAARDWRYAFFKSIMLKNNYSALVTAHHSD 123

Query: 122 DQAETVFMRILRGSRLRYLSGIKQVSAFANGQLIRPFLPYKKELLPNIFHFEDASNASSD 181
           DQAET+ MR++RGSRLR+LSGIK V  FANGQLIRPFL + K+ LP IFHFED+SN
Sbjct: 124 DQAETILMRLIRGSRLRHLSGIKSVQPFANGQLIRPFLTFSKKDLPEIFHFEDSSNRELS 183

Query: 182 YLRNRIRNVYFPALERENNQLKDSLITLSEETECLFTALTDLTRSIEVTNCYDFLRQTHS 241
           +LRNR+RN Y P L++EN +    L  L+ E   LF A  +LT  I  T+  +F  Q+ S
Sbjct: 184 FLRNRVRNNYLPLLKQENPRFIQGLNQLALENSLLFQAFKELTNHITTTDLTEFNEQSKS 243

Query: 242 VQEFLLQDYISKFPDLQVSKEQFRVILKLIRTKANIDYTIKSGYFLHKDYESFHITKIHP 301
           +Q FLLQDY+  FPDL + K QF  +L++I+T     Y +K  Y++  D  SF ITKI P
Sbjct: 244 IQYFLLQDYLEGFPDLDLKKSQFTQLLQIIQTAKQGYYYLKKDYYIFIDKFSFKITKIVP 303

Query: 302 KTDSFKVEKRLELHNIQIFSQYLFSY--GKFISQADITIPIYDTSPIILRRRKEGDRIFL 359
           KT+  K EK LE  +   +  Y FS+       Q  ++IP++  S I LR R+ GD I
Sbjct: 304 KTELVKEEKMLEYDSNLCYRDYYFSFMPKSNEDQGQVSIPLFSLSSIKLRSRQSGDYISF 363

Query: 360 GNHTKKIRRLFIDEKITLKEREEAVIGEQNKELIFVIVAGRTYLRKPSEHDIMKGKLYIE 419
           G+ +KKIRRLFIDEK T+ ER+ A+IGEQ++++IFV++  +TYLRK  +HDIM  KLYI+
Sbjct: 364 GHFSKKIRRLFIDEKFTIAERQNAIIGEQDEQIIFVLIGNKTYLRKACKHDIMLAKLYID 423

Query: 420 NLEK 423
            LEK
Sbjct: 424 KLEK 427
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1664

A DNA sequence (GBSx1759) was identified in *S.agalactiae* <SEQ ID 5159> which encodes the amino acid sequence <SEQ ID 5160>. This protein is predicted to be hypoxanthine-guanine phosphoribosyltransferase (hpt). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA48876 GB:X69123 hypoxanthine guanine
          phosphoribosyltransferase [Lactococcus lactis]
 Identities = 121/179 (67%), Positives = 152/179 (84%), Gaps = 1/179 (0%)

Query: 2   LENDIKKVLYSEEDIILKTKELGAKLTADYAGKNPLLVGVLKGSVPFMAELLKHIDTHVE 61
           L+  I+KVL SEE+II K+KELG  LT +Y GKNPL++G+L+GSVPF+AEL+KHID H+E
Sbjct: 6   LDKAIEKVLVSEEEIIEKSKELGEILTKEYEGKNPLVLGILRGSVPFLAELIKHIDCHLE 65
```

```
Query: 62  IDFMVVSSYHGGTTSSGEVKILKDVDTNIEGRDVIFIEDIIDTGRTLKYLRDMFKYRQAN 121
            DFM VSSYHGGT SSGEVK++ DVDT ++GRD++ +EDIIDTGRTLKYL+++ ++R AN
Sbjct: 66  TDFMTVSSYHGGTKSSGEVKLILDVDTAVKGRDILIVEDIIDTGRTLKYLKELLEHRGAN 125

Query: 122 SVKVATLFDKPEGRLVDIDADYVCYDIPNEFIVGFGLDYAENYRNLPYVGVLKEEIYSK 180
            VK+ TL DKPEGR+V+I  DY  + IPNEF+VGFGLDY ENYRNLPYVGVLK E+Y+K
Sbjct: 126 -VKIVTLLDKPEGRIVEIKPDYSGFTIPNEFVVGFGLDYEENYRNLPYVGVLKPEVYNK 183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5161> which encodes the amino acid sequence <SEQ ID 5162>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/180 (85%), Positives = 171/180 (95%)

Query: 1   MLENDIKKVLYSEEDIILKTKELGAKLTADYAGKNPLLVGVLKGSVPFMAELLKHIDTHV 60
           MLE DI+K+LYSE DII KTK+LG +LT DY  KNPL++GVLKGSVPFMAEL+KHIDTHV
Sbjct: 1   MLEQDIQKILYSENDIIRKTKKLGEQLTKDYQEKNPLMIGVLKGSVPFMAELMKHIDTHV 60

Query: 61  EIDFMVVSSYHGGTTSSGEVKILKDVDTNIEGRDVIFIEDIIDTGRTLKYLRDMFKYRQA 120
           EIDFMVVSSYHGGT+SSGEVKILKDVDTNIEGRD+I +EDIIDTGRTLKYLRDMFKYR+A
Sbjct: 61  EIDFMVVSSYHGGTSSSGEVKILKDVDTNIEGRDIIIVEDIIDTGRTLKYLRDMFKYRKA 120

Query: 121 NSVKVATLFDKPEGRLVDIDADYVCYDIPNEFIVGFGLDYAENYRNLPYVGVLKEEIYSK 180
           N++K+ATLFDKPEGR+V I+ADYVCY+IPNEFIVGFGLDYAENYRNLPYVGVLKEE+YSK
Sbjct: 121 NTIKIATLFDKPEGRVVKIEADYVCYNIPNEFIVGFGLDYAENYRNLPYVGVLKEEVYSK 180
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1665

A DNA sequence (GBSx1760) was identified in *S.agalactiae* <SEQ ID 5163> which encodes the amino acid sequence <SEQ ID 5164>. This protein is predicted to be cell division protein FtsH (ftsH). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11    Transmembrane  139 - 155 ( 133 - 158)
   INTEGRAL    Likelihood = -4.62    Transmembrane    8 -  24 (   7 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC16243 GB:AF061748 cell division protein FtsH [Streptococcus pneumoniae] (ver 2)
 Identities = 490/652 (75%), Positives = 561/652 (85%), Gaps = 5/652 (0%)

Query: 5   KNNGFLKNSFIYILLIIAVITTFQYYLKGTSSQ-NQQISYTKLVKQLKAGEIKSISYQPS 63
           +NNG +KN F+++L I  ++T FQY+  G +S  +QQI+YT+LV+++  G +K ++YQP+
Sbjct: 4   QNNGLIKNPFLWLLFIFFLVTGFQYFYSGNNSGGSQQINYTELVQEITDGNVKELTYQPN 63
```

```
Query: 64  GGVVEVSGTYKKAKTIKSANSFTFLGGSVATKVTGFNSVILPNDSSIKSLVSAAEENNTN 123
           G V+EVSG YK  KT K      F   SV TKV  F S ILP D+++  L   A ++
Sbjct: 64  GSVIEVSGVYKNPKTSKEGTGIQFFTPSV-TKVEKFTSTILPADTTVSELQKLATDHKAE 122

Query: 124 IQVKHESSSGTWISYIASFLPLVIMIGFFMMMMNQGGGGGARGAMSFGKNKARSSSKDEV 183
           + VKHESSSG WI+ + S +P  I+  F   MM   GGG  R  MSFG++KA++++K+++
Sbjct: 123 VTVKHESSSGIWINLLVSIVPFGILFFFLFSMMGNMGGGNGRNPMSFGRSKAKAANKEDI 182

Query: 184 KVRFSDVAGAEEEKQELIEVVDFLKDPKRYKSLGARIPAGVLLEGPPGTGKTLLAKAVAG 243
           KVRFSDVAGAEEEKQEL+EVV+FLKDPKR+  LGARIPAGVLLEGPPGTGKTLLAKAVAG
Sbjct: 183 KVRFSDVAGAEEEKQELVEVVEFLKDPKRFTKLGARIPAGVLLEGPPGTGKTLLAKAVAG 242

Query: 244 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGMGGG 303
           EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKA  AIIFIDEIDAVGR+RG G+GGG
Sbjct: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAAPAIIFIDEIDAVGRQRGVGLGGG 302

Query: 304 NDEREQTLNQLLIEMDGFEGNESIIVIAATNRSDVLDPALLRPGRFDRKVLVGQPDVKGR 363
           NDEREQTLNQLLIEMDGFEGNE IIVIAATNRSDVLDPALLRPGRFDRKVLVG+PDVKGR
Sbjct: 303 NDEREQTLNQLLIEMDGFEGNEGIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362

Query: 364 EAILRVHAKNKPLADNVDLKVVAQQTPGFVGADLENVLNEAALVAARRNKKVIDASDIDE 423
           EAIL+VHAKNKPLA++VDLK+VAQQTPGFVGADLENVLNEAALVAARRNK +IDASDIDE
Sbjct: 363 EAILKVHAKNKPLAEDVDLKLVAQQTPGFVGADLENVLNEAALVAARRNKSIIDASDIDE 422

Query: 424 AEDRVIAGPSKKDRTISERERAMVAYHEAGHTIVGLILSNARVVHKVTIVPRGRAGGYMI 483
           AEDRVIAGPSKKD+T+S++ER +VAYHEAGHTIVGL+LSNARVVHKVTIVPRGRAGGYMI
Sbjct: 423 AEDRVIAGPSKKDKTVSQKERELVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482

Query: 484 ALPKEDQMLLSKDDMKEQLAGLMGGRVAEEIIFNAQTTGASNDFEQATAMARAMVTEYGM 543
           ALPKEDQMLLSK+DMKEQLAGLMGGRVAEEIIFN QTTGASNDFEQAT MARAMVTEYGM
Sbjct: 483 ALPKEDQMLLSKEDMKEQLAGLMGGRVAEEIIFNVQTTGASNDFEQATQMARAMVTEYGM 542

Query: 544 SEKLGPVQYEGNHAMMAGQMSPEKSYSAQTAQLIDDEVRHLLNEARNKAADIINENRDTH 603
           SEKLGPVQYEGNHAM+ G  SP+KS S QTA  ID+EVR LLNEARNKAA+II  NR+TH
Sbjct: 543 SEKLGPVQYEGNHAML-GAQSPQKSISEQTAYEIDEEVRSLLNEARNKAAEIIQSNRETH 601

Query: 604 KLIAEALLKYETLDAAQIKSIFETGKMPETENDEDKARALSYDEIKEKMQEE 655
           KLIAEALLKYETLD+ QIK+++ETGKMPE    E+++ ALSYDE+K KM +E
Sbjct: 602 KLIAEALLKYETLDSTQIKALYETGKMPEAV--EEESHALSYDEVKSKMNDE 651
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5165> which encodes the amino acid sequence <SEQ ID 5166>. Analysis of this protein sequence reveals the following:

```
        Possible site: 38

   >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.38   Transmembrane  138 - 154 ( 132 - 158)

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC16243 GB:AF061748 cell division protein FtsH [Streptococcus pneumoniae] (ver 2)
 Identities = 487/654 (74%), Positives = 565/654 (85%), Gaps = 7/654 (1%)

Query: 5   KNNGFVKNSFIYILMIIVVITGFQFYLKGTSTQ-SQQISYSKLIKHLKAGDIKSLSYQPS 63
           +NNG +KN F+++L I  ++TGFQ++  G ++  SQQI+Y++L++ +  G++K L+YQP+
Sbjct: 4   QNNGLIKNPFLWLLFIFFLVTGFQYFYSGNNSGGSQQINYTELVQEITDGNVKELTYQPN 63

Query: 64  GSIIEVKGKYEKPQKVTVNSGLSFLGGRASTQVTEFSSLVLPSDTILKEMTAAADKNGTE 123
           GS+IEV G Y+ P+     +G+ F      T+V +F+S +LP+DT + E+   A  +  E
Sbjct: 64  GSVIEVSGVYKNPKTSKEGTGIQFFTPSV-TKVEKFTSTILPADTTVSELQKLATDHKAE 122

Query: 124 LTVKQESSSGTWITFLMSFLPIVIFAAFMMMMM-NQGGGGARGAMSFGKNKAKSQSKGNV 182
```

```
           +TVK ESSSG WI  L+S +P  I   F+  MM N GGG  R  MSFG++KAK+ +K ++
Sbjct: 123 VTVKHESSSGIWINLLVSIVPFGILFFFLFSMMGNMGGGNGRNPMSFGRSKAKAANKEDI 182

Query: 183 KVRFTDVAGAEEEKQELVEVVDFLKNPKKYKSLGARIPAGVLLEGPPGTGKTLLAKAVAG 242
           KVRF+DVAGAEEEKQELVEVV+FLK+PK++  LGARIPAGVLLEGPPGTGKTLLAKAVAG
Sbjct: 183 KVRFSDVAGAEEEKQELVEVVEFLKDPKRFTKLGARIPAGVLLEGPPGTGKTLLAKAVAG 242

Query: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGMGGG 302
           EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKA  AIIFIDEIDAVGR+RG G+GGG
Sbjct: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAAPAIIFIDEIDAVGRQRGVGLGGG 302

Query: 303 NDEREQTLNQLLIEMDGFEGNENIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362
           NDEREQTLNQLLIEMDGFEGNE IIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR
Sbjct: 303 NDEREQTLNQLLIEMDGFEGNEGIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362

Query: 363 EAILRVHAKNKPLANDVNLKVVAQQTPGFVGADLENVLNEAALVAARRNKIKIDASDIDE 422
           EAIL+VHAKNKPLA DV+LK+VAQQTPGFVGADLENVLNEAALVAARRNK  IDASDIDE
Sbjct: 363 EAILKVHAKNKPLAEDVDLKLVAQQTPGFVGADLENVLNEAALVAARRNKSIIDASDIDE 422

Query: 423 AEDRVIAGPSKKDRTISQKEREMVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482
           AEDRVIAGPSKKD+T+SQKERE+VAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI
Sbjct: 423 AEDRVIAGPSKKDKTVSQKERELVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482

Query: 483 ALPKEDQMLLSKEDLKEQLAGLMGGRVAEEIVFNAQTSGASNDFEQATQIARAMVTEYGM 542
           ALPKEDQMLLSKED+KEQLAGLMGGRVAEEI+FN QT+GASNDFEQATQ+ARAMVTEYGM
Sbjct: 483 ALPKEDQMLLSKEDMKEQLAGLMGGRVAEEIIFNVQTTGASNDFEQATQMARAMVTEYGM 542

Query: 543 SEKLGPVQYEGNHAMMPGQISPEKAYSAQTAQMIDDEVRELLNQARNQAADIINENRDTH 602
           SEKLGPVQYEGNHAM+  Q SP+K+ S QTA  ID+EVR LLN+ARN+AA+II  NR+TH
Sbjct: 543 SEKLGPVQYEGNHAMLGAQ-SPQKSISEQTAYEIDEEVRSLLNEARNKAAEIIQSNRETH 601

Query: 603 KLIAEALLKYETLDAAQIKSIYETGKMPVDLETDDNEAHALSYDEIKNKMTESE 656
           KLIAEALLKYETLD+ QIK++YETGKMP   E  + E+HALSYDE+K+KM + +
Sbjct: 602 KLIAEALLKYETLDSTQIKALYETGKMP---EAVEEESHALSYDEVKSKMNDEK 652
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 550/657 (83%), Positives = 612/657 (92%), Gaps = 2/657 (0%)

Query: 1   MKNNKNNGFLKNSFIYILLIIAVITTFQYYLKGTSSQNQQISYTKLVKQLKAGEIKSISY 60
           MKNNKNNGF+KNSFIYIL+II VIT FQ+YLKGTS+Q+QQISY+KL+K LKAG+IKS+SY
Sbjct: 1   MKNNKNNGFVKNSFIYILMIIVVITGFQFYLKGTSTQSQQISYSKLIKHLKAGDIKSLSY 60

Query: 61  QPSGGVVEVSGTYKKAKTIKSANSFTFLGGSVATKVTGFNSVILPNDSSIKSLVSAAEEN 120
           QPSG ++EV G Y+K + +   +  +FLGG  +T+VT F+S++LP+D+ +K + +AA++N
Sbjct: 61  QPSGSIIEVKGKYEKPQKVTVNSGLSFLGGRASTQVTEFSSLVLPSDTILKEMTAAADKN 120

Query: 121 NTNIQVKHESSSGTWISYIASFLPLVIMIGFFMMMMNQGGGGGARGAMSFGKNKARSSSK 180
            T + VK ESSSGTWI+++ SFLP+VI   F MMMMNQGGGG ARGAMSFGKNKA+S SK
Sbjct: 121 GTELTVKQESSSGTWITFLMSFLPIVIFAAFMMMMMNQGGGG-ARGAMSFGKNKAKSQSK 179

Query: 181 DEVKVRFSDVAGAEEEKQELIEVVDFLKDPKRYKSLGARIPAGVLLEGPPGTGKTLLAKA 240
             VKVRF+DVAGAEEEKQEL+EVVDFLK+PK+YKSLGARIPAGVLLEGPPGTGKTLLAKA
Sbjct: 180 GNVKVRFTDVAGAEEEKQELVEVVDFLKNPKKYKSLGARIPAGVLLEGPPGTGKTLLAKA 239

Query: 241 VAGEAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGM 300
           VAGEAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGM
Sbjct: 240 VAGEAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGM 299

Query: 301 GGGNDEREQTLNQLLIEMDGFEGNESIIVIAATNRSDVLDPALLRPGRFDRKVLVGQPDV 360
           GGGNDEREQTLNQLLIEMDGFEGNE+IIVIAATNRSDVLDPALLRPGRFDRKVLVG+PDV
Sbjct: 300 GGGNDEREQTLNQLLIEMDGFEGNENIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDV 359

Query: 361 KGREAILRVHAKNKPLADNVDLKVVAQQTPGFVGADLENVLNEAALVAARRNKKVIDASD 420
           KGREAILRVHAKNKPLA++V+LKVVAQQTPGFVGADLENVLNEAALVAARRNK  IDASD
Sbjct: 360 KGREAILRVHAKNKPLANDVNLKVVAQQTPGFVGADLENVLNEAALVAARRNKIKIDASD 419

Query: 421 IDEAEDRVIAGPSKKDRTISERERAMVAYHEAGHTIVGLILSNARVVHKVTIVPRGRAGG 480
```

```
              IDEAEDRVIAGPSKKDRTIS++ER MVAYHEAGHTIVGL+LSNARVVHKVTIVPRGRAGG
Sbjct: 420 IDEAEDRVIAGPSKKDRTISQKEREMVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGG 479

Query: 481 YMIALPKEDQMLLSKDDMKEQLAGLMGGRVAEEIIFNAQTTGASNDFEQATAMARAMVTE 540
           YMIALPKEDQMLLSK+D+KEQLAGLMGGRVAEEI+FNAQT+GASNDFEQAT +ARAMVTE
Sbjct: 480 YMIALPKEDQMLLSKEDLKEQLAGLMGGRVAEEIVFNAQTSGASNDFEQATQIARAMVTE 539

Query: 541 YGMSEKLGPVQYEGNHAMMAGQMSPEKSYSAQTAQLIDDEVRHLLNEARNKAADIINENR 600
           YGMSEKLGPVQYEGNHAMM GQ+SPEK+YSAQTAQ+IDDEVR LLN+ARN+AADIINENR
Sbjct: 540 YGMSEKLGPVQYEGNHAMMPGQISPEKAYSAQTAQMIDDEVRELLNQARNQAADIINENR 599

Query: 601 DTHKLIAEALLKYETLDAAQIKSIFETGKMP-ETENDEDKARALSYDEIKEKMQEED 656
           DTHKLIAEALLKYETLDAAQIKSI+ETGKMP + E D+++A ALSYDEIK KM E +
Sbjct: 600 DTHKLIAEALLKYETLDAAQIKSIYETGKMPVDLETDDNEAHALSYDEIKNKMTESE 656
```

SEQ ID 5164 (GBS115) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 8; MW 73kDa) and in Figure 39 (lane 3; MW 73.3kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1666

A DNA sequence (GBSx1769) was identified in *S.agalactiae* <SEQ ID 5167> which encodes the amino acid sequence <SEQ ID 5168>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2983(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1667

A DNA sequence (GBSx1770) was identified in *S.agalactiae* <SEQ ID 5169> which encodes the amino acid sequence <SEQ ID 5170>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2424(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9547> which encodes amino acid sequence <SEQ ID 9548> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12187 GB:Z99106 similar to homoserine dehydrogenase [Bacillus subtilis]
 Identities = 223/448 (49%), Positives = 313/448 (69%)
```

```
Query: 1   MKVVKFGGSSLASSQQLYKVLNIIKSDYTRRFVVVSAPGKRYEEDLKMTDALIQYYQNYI 60
           MKVVKFGGSSLAS  QL KV +I+ SD  R+ VVVSAPGK Y ED K+TD LI   + Y+
Sbjct: 1   MKVVKFGGSSLASGAQLDKVFHIVTSDPARKAVVVSAPGKHYAEDTKVTDLLIACAEQYL 60

Query: 61  NGKDIVKDQTWIINRYQEIISDLSLGSTIAEEITRSIEQLASLPIENNQFLYDCFLAAGE 120
                 +    ++ RY  I ++L LG +I E+I   + L      N +   D   A+GE
Sbjct: 61  ATGSAPELAEAVVERYALIANELQLGQSIIEKIRDDLFTLLEGDKSNPEQYLDAVKASGE 120

Query: 121 DNNAKLVATFFNQNDIPARYVHPNEAGIIVTKEPCNARIIPGSYDKIENLCLYNEVLVIP 180
           DNNAKL+A +F    + A YV+P +AG+ VT EP NA+++P SY  +  L   + +++ P
Sbjct: 121 DNNAKLIAAYFRYKGVKAEYVNPKDAGLFVTNEPGNAQVLPESYQNLYRLRERDGLIIFP 180

Query: 181 GFFGVTEDNQICTFSRGGSDITGSLIAAGIKADLYENFTDVDGIFAAHPGVVKNPHAIPE 240
           GFFG ++D  + TFSR GSDITGS++A G++ADLYENFTDVD +++ +P  V+NP  I E
Sbjct: 181 GFFGFSKDGDVITFSRSGSDITGSILANGLQADLYENFTDVDAVYSVNPSFVENPKEISE 240

Query: 241 LTYKEMRELAYAGFSVLHDEALLPAYRGRIPLVIKNTNNPQQPGTKIVLKHTRSNIAVTG 300
           LTY+EMREL+YAGFSV HDEAL+PA+R  IP+ IKNTNNP   GT++V K   +N  V G
Sbjct: 241 LTYREMRELSYAGFSVFHDEALIPAFRAGIPVQIKNTNNPSAEGTRVVSKRDNTNGPVVG 300

Query: 301 IASDSRFASINVSKYLMNREVGFGRKVLQILEDLNISFEHMPTGIDDLSIVLREKELTPI 360
           IASD+ F SI +SKYLMNRE+GFGR+ LQILE+  +++EH+P+GIDD++I+LR+ ++
Sbjct: 301 IASDTGFCSIYISKYLMNREIGFGRRALQILEEHGLTYEHVPSGIDDMTIILRQGQMDAA 360

Query: 361 KEQEILNYLTRKLEVDYVDIQHNLSTIVIVGENMKSQIGVTATATQALSREKINITMISQ 420
            E+ ++  +   L  D V ++H+L+ I++VGE M+  +G TA A +ALS  ++NI MI+Q
Sbjct: 361 TERSVIKRIEEDLHADEVIVEHHLALIMVVGEAMRHNVGTTARAAKALSEAQVNIEMINQ 420

Query: 421 GSSEVSIMFVINSKDEKRAIKALYETFF 448
           GSSEVS+MF +   +E++A++ALY+ FF
Sbjct: 421 GSSEVSMMFGVKEAEERKAVQALYQEFF 448
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1668

A DNA sequence (GBSx1771) was identified in *S.agalactiae* <SEQ ID 5171> which encodes the amino acid sequence <SEQ ID 5172>. This protein is predicted to be CbbY family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2699(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF96016 GB:AE004353 CbbY family protein [Vibrio cholerae]
 Identities = 59/190 (31%), Positives = 93/190 (48%), Gaps = 10/190 (5%)

Query: 4   YKAIIFDMDGVLFDTELFYYKRRERFLKQHGITIDHLPMNFFIGGNMKQVWKSVLGDQYD 63
           ++A IFDMDG+L DTE    +  +      G+          IG N K +   +L   Y
Sbjct: 6   FQAAIFDMDGLLLDTERVCMRVFQEACTACGLPFRQEVYLSVIGCNAKTI-NGILSQAYG 64

Query: 64  TWDIDKL----QQDYSRYKEDNPLPYKDLIFQDCKRVIEKLHHKGYLLGLASSSTRHDIM 119
             D+ +L    +Q Y+       +P+KD +      ++E L  +   + +A+S+ +   +
Sbjct: 65  E-DLPRLHNEWRQRYNAVVMHEAIPHKDGVIA----LLEWLKARSIPVAVATSTQKEVAL 119

Query: 120 LALESFNLDTYFKVILSGEEFSESKPNPAIYNRAAELLDIPKQQILIVEDSEKGITAGIA 179
```

```
           + L+   LD YF  I +G E ++ KP+P IY  AAE L +  QQ L  EDS  GI A +A
Sbjct: 120 IKLQLAGLDHYFANITTGCEVTQGKPHPEIYLLAAERLGVEPQQCLAFEDSNNGIKAAMA 179

Query: 180 AGIDVWAIED 189
           A +  + I D
Sbjct: 180 AQMHAFQIPD 189
```

There is also homology to SEQ ID 448.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1669

A DNA sequence (GBSx1772) was identified in *S.agalactiae* <SEQ ID 5173> which encodes the amino acid sequence <SEQ ID 5174>. This protein is predicted to be Pseudomonas putida enoyl-CoA hydratase II homologue (b1394). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18    Transmembrane  128 - 144 ( 128 - 145)
   INTEGRAL    Likelihood = -1.06    Transmembrane  154 - 170 ( 154 - 170)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9549> which encodes amino acid sequence <SEQ ID 9550> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5175> which encodes the amino acid sequence <SEQ ID 5176>. Analysis of this protein sequence reveals the following:

```
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.08    Transmembrane  110 - 126 ( 109 - 128)

----- Final Results -----
               bacterial membrane --- Certainty=0.2232(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 150/263 (57%), Positives = 197/263 (74%)

Query: 19  LKFENIIYGIDGNVATIMLNRPDISNGFNIPMCQEIIDAIRLVSENKDVMFLVIEAQGPI 78
            ++F++II+ +  ++AT+ LNRP++SNGFNIP+CQEI+ A+  V  +  V FL+I+A G +
Sbjct: 1   MQFKHIIFDVVDDLATLTLNRPEVSNGFNIPICQEILVALAEVKRDTSVRFLLIKAVGKV 60

Query: 79  FSIGGDLKVMKAAVESDDISSLTKIAELVNQISYDLLQLEKPVVMCVDGAVAGAAANIAL 138
            FS+GGDL  M+ AV  D++ SL KIAELV +IS+ +  L KPV++C DGAVAGAA NIAL
Sbjct: 61  FSVGGDLVEMQEAVAKDNVQSLVKIAELVQEISFAIKHLPKPVILCADGAVAGAAFNIAL 120

Query: 139 AADFVIASKKSKFIQAFVGVGLAPDAGGLLLLSKSIGITRAVQLALTGESLSAEKAEALG 198
            A DF IAS ++KFIQAFV VGLAPDAGGL LL++++G+ RA  L +TGE ++A+K    G
Sbjct: 121 AVDFCIASTQTKFIQAFVNVGLAPDAGGLFLLTRAVGLNRATHLVMTGEGITADKGLDYG 180

Query: 199 IVYKLCESDKIGKIKDQLLKRLSRHSINSYQAIKSLAWEAAFKDWEQYKKLELQLQESLA 258
             VY+  ESDK+ K+  QLLKRL R S NSY  +KSL W++ F  WE Y K EL +QE LA
Sbjct: 181 FVYRTAESDKLDKVCLQLLKRLRRGSSNSYAGMKSLVWQSFFTGWEDYAKAELAIQEELA 240
```

```
Query: 259 FKQDFKEGVRAHADRRRPNFLGK 281
           FK+DFKEGV A  +RRRPNF GK
Sbjct: 241 FKEDFKEGVIAFGERRRPNFQGK 263
```

A related GBS gene <SEQ ID 8877> and protein <SEQ ID 8878> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
SRCFLG: 0
McG: Length  of  UR:   9
     Peak Value of UR:   1.45
     Net Charge of CR: -1
McG: Discrim Score:     -5.99
GvH: Signal Score (-7.5): -4.37
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value:  -2.18 threshold:  0.0
   INTEGRAL    Likelihood = -2.18   Transmembrane  110 - 126 ( 110 - 127)
   INTEGRAL    Likelihood = -1.06   Transmembrane  136 - 152 ( 136 - 152)
   PERIPHERAL  Likelihood =  1.32      49
 modified ALOM score:   0.94
icm1 HYPID: 7  CFP: 0.187

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01047(355 - 1143 of 1443)
GP|3253198|gb|AAC24330.1||AF029714(1 - 263 of 263) PhaB {Pseudomonas putida}
%Match = 15.4
%Identity = 33.3  %Similarity = 56.4
Matches = 88  Mismatches = 113  Conservative Sub.s = 61

96        126       156       186       216       246       276       306
*KTVRRGLQLVLQPVLMCGLLKINTLE*ISRRLMY**AI*VNFL*N*ITIKNGKFNSVFLFFILP*KLGL**NTKHDNLI

336       366       396       426       456       486       516       546
IKLFFIFLSLLKRGDILKFENIIYGIDGNVATIMLNRPDISNGFNIPMCQEIIDAIRLVSENKDVMFLVIEAQGPIFSIG
                : |::|:: |:  || : ||||:  | ||  |  |: :|:: | :: |   |:: |:|  |  |
                MTFQHILFSIEDGVAFLSLNRPEQLNSFNAAMHLEVREALKQVRQSSDARVLLLTAEGRGFCAG
                        10        20        30        40        50        60

576       606       636       666       696       726       756       786
GDLKVMKAAVESDDISSLTKIAELVNQISYDLLQLEKPVVMCVDGAVAGAAANIALAADFVIASKKSKFIQAFVGVGLAP
 ||     | :::       | :: | :   |  |  ||:  |:|  ||| ||| || |:|:| : : |||||  :|| |
QDLSDRNVAPDAEVPDLGESIDKFYNPLVRTLRDLPLPVICAVNGVAAGAGANIPLACDLVLAGRSASFIQAFCKIGLVP
              80        90       100       110       120       130       140

816       846       876       906       936       966       996       1026
DAGGLLLLSKSIGITRAVQLALTGESLSAEKAEALGIVYKLCESDKIGKIKDQLLKRLSRHSINSYQAIKSLAWEAAFKD
|:||  || : :|: ||  ||: || | ||:|:  |::::: :   :      | ::|: :       ||  :  |:| :
DSGGTWLLPRLVGMARAKALAMLGERLGAEQAQQWGLIHRVVDDAALRDEALTLARQLASQPTYGLALIK-RSLNASFDN
             160       170       180       190       200       210       220

1053      1083      1113      1143      1173      1203      1233      1263
-WEQYKKLELQLQESLAFKQDFKEGVRAHADRRRPNFLGK*FENQII*D*SLANKFEL*YNLIIKV*CEVVISWNTIRLI
 :::  :||  ||      :|::||| |  ::| | | |:
GFDEQLELERDLQRLAGRSEDYREGVSAFMNKRTPAFKGR
              240       250       260
```

SEQ ID 8878 (GBS374) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 8; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 2; MW 57kDa).

The GBS374-GST fusion product was purified (Figure 215, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 307), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1670

A DNA sequence (GBSx1773) was identified in *S.agalactiae* <SEQ ID 5177> which encodes the amino acid sequence <SEQ ID 5178>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD05186 GB:AF110185 unknown [Burkholderia pseudomallei]
 Identities = 30/102 (29%), Positives = 60/102 (58%)

Query: 32  DVSLKEMHTIEIIGKHSEVTPSDVARELMLTLGTVTTSLNKLEKKGYIERKRSSIDRRVV 91
            +++ +++  I ++ +    TP +++R+L    G++T  L++LEKKG++ R RS  DRRV+
Sbjct: 39  ELTAQQISVILLLARGYARTPFELSRKLSYDSGSMTRMLDRLEKKGFVVRARSESDRRVI 98

Query: 92  HLSLTKRGRLLDRLHSKFHKSMVSHIIEDLGEEDIKMLTSAL 133
             L+LT+RG    R       + ++  +E    +++ +LT  L
Sbjct: 99  ELALTERGAHAARALPALIATELNAQLEGFSADELALLTDLL 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5179> which encodes the amino acid sequence <SEQ ID 5180>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1412(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 111/144 (77%), Positives = 129/144 (89%)

Query: 1   MEYDQINSYLVDIFNRIMIIEEMSLKTSQFSDVSLKEMHTIEIIGKHSEVTPSDVARELM 60
            +EYD+I  YLVDIFNRI++IEEMSLKTSQFSDVSLKEMHTIEIIGK+ +VTPSD+ARELM
Sbjct: 7   LEYDKIYPYLVDIFNRILVIEEMSLKTSQFSDVSLKEMHTIEIIGKYDQVTPSDIARELM 66

Query: 61  LTLGTVTTSLNKLEKKGYIERKRSSIDRRVVHLSLTKRGRLLDRLHSKFHKSMVSHIIED 120
            +TLGTVTTSLNKLE KGYI R RS  DRRVV+LSLTKRGRLLDRLH+KFHK+MV H+I D
Sbjct: 67  VTLGTVTTSLNKLEAKGYIARTRSRSDRRVVYLSLTKRGRLLDRLHAKFHKNMVGHVIAD 126
```

```
Query: 121 LGEEDIKMLTSALGNLHKFLEDLV 144
           + +E+++ L    LGNLH+FLEDLV
Sbjct: 127 MSDEEMQALVRGLGNLHQFLEDLV 150
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1671

A DNA sequence (GBSx1774) was identified in *S.agalactiae* <SEQ ID 5181> which encodes the amino acid sequence <SEQ ID 5182>. This protein is predicted to be 3-oxoacyl-(acyl-carrier-protein) synthase III (fabH-2). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  103 - 119 ( 103 - 119)

----- Final Results -----
               bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98271 GB:AF197933 beta-ketoacyl-ACP synthase III
           [Streptococcus pneumoniae]
 Identities = 225/324 (69%), Positives = 276/324 (84%), Gaps = 1/324 (0%)

Query: 1   MVFAKISQLAHYAPSQIIKNEDLSLIMDTSDDWISSRTGIKQRHISKNETTADLANKVAE 60
           M FAKISQ+AHY P Q++ N DL+ IMDT+D+WISSRTGI+QRHIS+ E+T+DLA +VA+
Sbjct: 1   MAFAKISQVAHYVPEQVVTNHDLAQIMDTNDEWISSRTGIRQRHISRTESTSDLATEVAK 60

Query: 61  QLIEKSGYSASQIDFIIVATMTPDSMMPSTAARVQAHIGASNAFAFDLSAACSGFVFALS 120
           +L+ K+G +  ++DFII+AT+TPDSMMPSTAARVQA+IGA+ AFAFDL+AACSGFVFALS
Sbjct: 61  KLMAKAGITGEELDFIILATITPDSMMPSTAARVQANIGANKAFAFDLTAACSGFVFALS 120

Query: 121 TAEKLISSGSYQKGLVIGAETVSKVLDWTDRGTAVLFGDGAGGVLLEASKEKHFLAESLN 180
           TAEK I+SG +QKGLVIG+ET+SK +DW+DR TAVLFGDGAGGVLLEAS+++HFLAESLN
Sbjct: 121 TAEKFIASGRFQKGLVIGSETLSKAVDWSDRSTAVLFGDGAGGVLLEASEQEHFLAESLN 180

Query: 181 TDGSR-QGLQSSQVGLNSPFSDEVLDDKFLKMDGRAIFDFAIKEVSKSINHLIETSYLEK 239
           +DGSR + L     GL+SPFSD+   D FLKMDGR +FDFAI++V+KSI   I+ S +E
Sbjct: 181 SDGSRSECLTYGHSGLHSPFSDQESADSFLKMDGRTVFDFAIRDVAKSIKQTIDESPIEV 240

Query: 240 EDIDYLFLHQANRRILDKMSRKIDIARDKFPENMMDYGNTSAASIPILLSESYENGLLKL 299
            D+DYL LHQAN RILDKM+RKI + R K P NMM+YGNTSAASIPILLSE  E GL+ L
Sbjct: 241 TDLDYLLLHQANDRILDKMARKIGVDRAKLPANMMEYGNTSAASIPILLSECVEQGLIPL 300

Query: 300 DGNQTILLSGFGGGLTWGSLIVKI 323
           DG+QT+LLSGFGGGLTWG+LI+ I
Sbjct: 301 DGSQTVLLSGFGGGLTWGTLILTI 324
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5183> which encodes the amino acid sequence <SEQ ID 5184>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  103 - 119 ( 103 - 120)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF98271 GB:AF197933 beta-ketoacyl-ACP synthase III
           [Streptococcus pneumoniae]
 Identities = 212/324 (65%), Positives = 263/324 (80%)

Query: 1   MIFSKISQVAHYVPQQLVTNNDLASIMDTSHEWIFSRTGIAERHISRDEMTSDLAIQVAD 60
           M F+KISQVAHYVP+Q+VTN+DLA IMDT+ EWI SRTGI +RHISR E TSDLA +VA
Sbjct: 1   MAFAKISQVAHYVPEQVVTNHDLAQIMDTNDEWISSRTGIRQRHISRTESTSDLATEVAK 60

Query: 61  QLLTQSGLKADAIDFIIVATISPDATMPSTAAKVQAAIAATSAFAFDMTAACSGFVFALA 120
           +L+ ++G+  + +DFII+ATI+PD+ MPSTAA+VQA I A  AFAFD+TAACSGFVFAL+
Sbjct: 61  KLMAKAGITGEELDFIILATITPDSMMPSTAARVQANIGANKAFAFDLTAACSGFVFALS 120

Query: 121 MADKLIASGAYQNGMVIGAETLSKLVNWQDRATAVLFGDGAGGVLLEASKDKHVLAETLH 180
            A+K IASG +Q G+VIG+ETLSK V+W DR+TAVLFGDGAGGVLLEAS+ +H LAE+L+
Sbjct: 121 TAEKFIASGRFQKGLVIGSETLSKAVDWSDRSTAVLFGDGAGGVLLEASEQEHFLAESLN 180

Query: 181 TDGARCQSLISGETSLSSPYSIGKKAIATIQMDGRAIFDFAIRDVSKSILTLMAQSDITK 240
           +DG+R + L  G + L SP+S  + A + ++MDGR +FDFAIRDV+KSI   + +S I
Sbjct: 181 SDGSRSECLTYGHSGLHSPFSDQESADSFLKMDGRTVFDFAIRDVAKSIKQTIDESPIEV 240

Query: 241 DDIDYCLLHQANRRILDKIARKIDVPREKFLENMMRYGNTSAASIPILLSEAVQKGQIRL 300
            D+DY LLHQAN RILDK+ARKI V R K   NMM YGNTSAASIPILLSE V++G I L
Sbjct: 241 TDLDYLLLHQANDRILDKMARKIGVDRAKLPANMMEYGNTSAASIPILLSECVEQGLIPL 300

Query: 301 DGTQKILLSGFGGGLTWGSLIVRI 324
           DG+Q +LLSGFGGGLTWG+LI+ I
Sbjct: 301 DGSQTVLLSGFGGGLTWGTLILTI 324
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 216/324 (66%), Positives = 271/324 (82%), Gaps = 1/324 (0%)

Query: 1   MVFAKISQLAHYAPSQIIKNEDLSLIMDTSDDWISSRTGIKQRHISKNETTADLANKVAE 60
           M+F+KISQ+AHY P Q++ N DL+ IMDTS +WI SRTGI +RHIS++E T+DLA +VA+
Sbjct: 1   MIFSKISQVAHYVPQQLVTNNDLASIMDTSHEWIFSRTGIAERHISRDEMTSDLAIQVAD 60

Query: 61  QLIEKSGYSASQIDFIIVATMTPDSMMPSTAARVQAHIGASNAFAFDLSAACSGFVFALS 120
           QL+ +SG  A  IDFIIVAT++PD+ MPSTAA+VQA I A++AFAFD++AACSGFVFAL+
Sbjct: 61  QLLTQSGLKADAIDFIIVATISPDATMPSTAAKVQAAIAATSAFAFDMTAACSGFVFALA 120

Query: 121 TAEKLISSGSYQKGLVIGAETVSKVLDWTDRGTAVLFGDGAGGVLLEASKEKHFLAESLN 180
            A+KLI+SG+YQ G+VIGAET+SK+++W DR TAVLFGDGAGGVLLEASK+KH LAE+L+
Sbjct: 121 MADKLIASGAYQNGMVIGAETLSKLVNWQDRATAVLFGDGAGGVLLEASKDKHVLAETLH 180

Query: 181 TDGSR-QGLQSSQVGLNSPFSDEVLDDKFLKMDGRAIFDFAIKEVSKSINHLIETSYLEK 239
           TDG+R Q L S +  L+SP+S        ++MDGRAIFDFAI++VSKSI  L+  S + K
Sbjct: 181 TDGARCQSLISGETSLSSPYSIGKKAIATIQMDGRAIFDFAIRDVSKSILTLMAQSDITK 240

Query: 240 EDIDYLFLHQANRRILDKMSRKIDIARDKFPENMMDYGNTSAASIPILLSESYENGLLKL 299
           +DIDY  LHQANRRILDK++RKID+ R+KF ENMM YGNTSAASIPILLSE+ + G ++L
Sbjct: 241 DDIDYCLLHQANRRILDKIARKIDVPREKFLENMMRYGNTSAASIPILLSEAVQKGQIRL 300

Query: 300 DGNQTILLSGFGGGLTWGSLIVKI 323
           DG Q ILLSGFGGGLTWGSLIV+I
Sbjct: 301 DGTQKILLSGFGGGLTWGSLIVRI 324
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1672

A DNA sequence (GBSx1775) was identified in *S.agalactiae* <SEQ ID 5185> which encodes the amino acid sequence <SEQ ID 5186>. This protein is predicted to be acyl carrier protein (acpP). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3083(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9551> which encodes amino acid sequence <SEQ ID 9552> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98272 GB:AF197933 acyl carrier protein [Streptococcus pneumoniae]
 Identities = 64/74 (86%), Positives = 67/74 (90%)

Query: 17 MAVFEKVQEIIVEELGKDAEEVTLNTTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN 76
           MAVFEKVQEIIVEELGKDA EVTL +TFDDLDADSLD+FQVISEIEDAFDIQIE E  L
Sbjct: 1  MAVFEKVQEIIVEELGKDASEVTLESTFDDLDADSLDLFQVISEIEDAFDIQIEAENDLK 60

Query: 77 TVGDLVAYVEEKVK 90
           TVGDLVAYVEE+ K
Sbjct: 61 TVGDLVAYVEEQAK 74
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5187> which encodes the amino acid sequence <SEQ ID 5188>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2995(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 70/74 (94%), Positives = 71/74 (95%)

Query: 17 MAVFEKVQEIIVEELGKDAEEVTLNTTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN 76
           MAVFEKVQEIIVEELGK+ EEVTL TTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN
Sbjct: 1  MAVFEKVQEIIVEELGKETEEVTLETTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN 60

Query: 77 TVGDLVAYVEEKVK 90
           TVGDLVAYVEEK K
Sbjct: 61 TVGDLVAYVEEKSK 74
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1673

A DNA sequence (GBSx1777) was identified in *S.agalactiae* <SEQ ID 5189> which encodes the amino acid sequence <SEQ ID 5190>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
```

```
    INTEGRAL    Likelihood = -0.27    Transmembrane  156 - 172 ( 156 - 173)

 ----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98273 GB:AF197933 trans-2-enoyl-ACP reductase II
          [Streptococcus pneumoniae]
 Identities = 257/318 (80%), Positives = 277/318 (86%), Gaps = 1/318 (0%)

Query: 1   MKTRITELLNIKYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60
           MKTRITELL I YPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS
Sbjct: 1   MKTRITELLKIDYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60

Query: 61  MTDKPFGVNIMLLSPFVDDIVDLVIEEGVKVVTTGAGNPGKYMERFHEAGITVIPVVPSV 120
           +TDKPFGVNIMLLSPFV+DIVDLVIEEGVKVVTTGAGNP KYMERFHEAGI VIPVVPSV
Sbjct: 61  LTDKPFGVNIMLLSPFVEDIVDLVIEEGVKVVTTGAGNPSKYMERFHEAGIIVIPVVPSV 120

Query: 121 ALAKRMEKLGADAIITEGMEAGGHIGKLTTMTLVRQVVDAVTIPVIAAGGIADGRGAAAG 180
           ALAKRMEK+GADA+I EGMEAGGHIGKLTTMTLVRQV  A++IPVIAAGGIADG GAAAG
Sbjct: 121 ALAKRMEKIGADAVIAEGMEAGGHIGKLTTMTLVRQVATAISIPVIAAGGIADGEGAAAG 180

Query: 181 FMLGADAVQVGTRFVVAKESNAHPNYKAKILKAKDIDTAVSAQVVGHPVRALKNKLVTTY 240
           FMLGA+AVQVGTRFVVAKESNAHPNYK KILKA+DIDT +SAQ  GH VRA+KN+L   +
Sbjct: 181 FMLGAEAVQVGTRFVVAKESNAHPNYKEKILKARDIDTTISAQHFGHAVRAIKNQLTRDF 240

Query: 241 SQAEKDYLAGRISINEI-EELGAGALRNAVVDGDVINGSVMAGQIAGLIKSEETCQEILE 299
             AEKD         EI E++GAGAL  AVV GDV  GSVMAGQIAGL+  EET +EIL+
Sbjct: 241 ELAEKDAFKQEDPDLEIFEQMGAGALAKAVVHGDVDGGSVMAGQIAGLVSKEETAEEILK 300

Query: 300 DIYSGARQVILSEASRWS 317
           D+Y GA + I  EASRW+
Sbjct: 301 DLYYGAAKKIQEEASRWT 318
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5191> which encodes the amino acid sequence <SEQ ID 5192>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.70    Transmembrane  106 - 122 ( 106 - 124)
    INTEGRAL    Likelihood = -0.22    Transmembrane  156 - 172 ( 156 - 173)

 ----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF98273 GB:AF197933 trans-2-enoyl-ACP reductase II
          [Streptococcus pneumoniae]
 Identities = 252/320 (78%), Positives = 276/320 (85%), Gaps = 1/320 (0%)

Query: 1   MKTRITELLNIDYPIFQGGMAWVADGDLAGAVSNAGGLGIIGGGNAPKEVVKANIDRVKA 60
           MKTRITELL IDYPIFQGGMAWVADGDLAGAVS AGGLGIIGGGNAPKEVVKANID++K+
Sbjct: 1   MKTRITELLKIDYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60

Query: 61  ITDRPFGVNIMLLSPFADDIVDLVIEEGVKVVTTGAGNPGKYMERLHQAGIIVVPVVPSV 120
           +TD+PFGVNIMLLSPF +DIVDLVIEEGVKVVTTGAGNP KYMER H+AGIIV+PVVPSV
Sbjct: 61  LTDKPFGVNIMLLSPFVEDIVDLVIEEGVKVVTTGAGNPSKYMERFHEAGIIVIPVVPSV 120

Query: 121 ALAKRMEKLGVDAVIAEGMEAGGHIGKLTTMSLVRQVVEAVSIPVIAAGGIADGHGAAAA 180
           ALAKRMEK+G DAVIAEGMEAGGHIGKLTTM+LVRQV  A+SIPVIAAGGIADG GAAA
Sbjct: 121 ALAKRMEKIGADAVIAEGMEAGGHIGKLTTMTLVRQVATAISIPVIAAGGIADGEGAAAG 180
```

```
Query: 181 FMLGAEAVQIGTRFVVAKESNAHQNFKDKILAAKDIDTVISAQVVGHPVRSIKNKLTSAY 240
           FMLGAEAVQ+GTRFVVAKESNAH N+K+KIL A+DIDT ISAQ  GH VR+IKN+LT  +
Sbjct: 181 FMLGAEAVQVGTRFVVAKESNAHPNYKEKILKARDIDTTISAQHFGHAVRAIKNQLTRDF 240

Query: 241 AKAEK-AFLIGQKTATDIEEMGAGSLRHAVIEGDVVNGSVMAGQIAGLVRKEESCETILK 299
             AEK AF          E+MGAG+L  AV+ GDV  GSVMAGQIAGLV KEE+ E ILK
Sbjct: 241 ELAEKDAFKQEDPDLEIFEQMGAGALAKAVVHGDVDGGSVMAGQIAGLVSKEETAEEILK 300

Query: 300 DIYYGAARVIQNEAKRWQSV 319
           D+YYGAA+ IQ EA RW  V
Sbjct: 301 DLYYGAAKKIQEEASRWTGV 320
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 253/319 (79%), Positives = 291/319 (90%)

Query: 1   MKTRITELLNIKYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60
           MKTRITELLNI YPIFQGGMAWVADGDLAGAVS AGGLGIIGGGNAPKEVVKANID++K+
Sbjct: 1   MKTRITELLNIDYPIFQGGMAWVADGDLAGAVSNAGGLGIIGGGNAPKEVVKANIDRVKA 60

Query: 61  MTDKPFGVNIMLLSPFVDDIVDLVIEEGVKVVTTGAGNPGKYMERFHEAGITVIPVVPSV 120
           +TD+PFGVNIMLLSPF DDIVDLVIEEGVKVVTTGAGNPGKYMER H+AGI V+PVVPSV
Sbjct: 61  ITDRPFGVNIMLLSPFADDIVDLVIEEGVKVVTTGAGNPGKYMERLHQAGIIVVPVVPSV 120

Query: 121 ALAKRMEKLGADAIITEGMEAGGHIGKLTTMTLVRQVVDAVTIPVIAAGGIADGRGAAAG 180
           ALAKRMEKLG DA+I EGMEAGGHIGKLTTM+LVRQVV+AV+IPVIAAGGIADG GAAA
Sbjct: 121 ALAKRMEKLGVDAVIAEGMEAGGHIGKLTTMSLVRQVVEAVSIPVIAAGGIADGHGAAAA 180

Query: 181 FMLGADAVQVGTRFVVAKESNAHPNYKAKILKAKDIDTAVSAQVVGHPVRALKNKLVTTY 240
           FMLGA+AVQ+GTRFVVAKESNAH N+K KIL AKDIDT +SAQVVGHPVR++KNKL + Y
Sbjct: 181 FMLGAEAVQIGTRFVVAKESNAHQNFKDKILAAKDIDTVISAQVVGHPVRSIKNKLTSAY 240

Query: 241 SQAEKDYLAGRISINEIEELGAGALRNAVVDGDVINGSVMAGQIAGLIKSEETCQEILED 300
           ++AEK +L G+ +  +IEE+GAG+LR+AV++GDV+NGSVMAGQIAGL++ EE+C+ IL+D
Sbjct: 241 AKAEKAFLIGQKTATDIEEMGAGSLRHAVIEGDVVNGSVMAGQIAGLVRKEESCETILKD 300

Query: 301 IYSGARQVILSEASRWSDL 319
           IY GA +VI +EA RW  +
Sbjct: 301 IYYGAARVIQNEAKRWQSV 319
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1674

A DNA sequence (GBSx1778) was identified in *S.agalactiae* <SEQ ID 5193> which encodes the amino acid sequence <SEQ ID 5194>. This protein is predicted to be MCAT (fabD). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1276(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with a *S.pneumoniae* sequence:

```
Identities = 203/306 (66%), Positives = 242/306 (78%), Gaps = 1/306 (0%)

Query: 1   MNKVSFLFAGQGAQKLGMARDLYETFPIVKETFDKASHVLGYDLRELIDKDLDKLNQTKY 60
           M K +FLFAGQGAQ LGM RD Y+ +PIVKET D+AS VLGYDLR LID + DKLNQT+Y
Sbjct: 1   MTKTAFLFAGQGAQYLGMGRDFYDQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60
```

```
Query: 61  TQPAILTTSTAIYRLILKEIELRPDMVAGLSLGEYSALVASGAIRFEDAVVLVARRGQLM 120
           TQPAIL TS AIYRL L+E   +PDMVAGLSLGEYSALVASGA+ FEDAV LVA+RG  M
Sbjct: 61  TQPAILATSVAIYRL-LQEKGYQPDMVAGLSLGEYSALVASGALDFEDAVALVAKRGAYM 119

Query: 121 EAAAPAGSGKMVAVLNADRQIIEDACKKASQFGIVSPANYNTPKQIVIGGESIAVNAAVE 180
           E AAPA SGKMVAVLN   ++IE+AC+KAS+ G+V+PANYNTP QIVI GE +AV+ AVE
Sbjct: 120 EEAAPADSGKMVAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAGEVVAVDRAVE 179

Query: 181 ELKQQGVKRLIPLNVSGPFHTALLKPASQKLSDVLDKVHFSVSEIPVIGNTEAQIMKKDD 240
            L++ G KRLIPL VSGPFHTALL+PASQKL++ L +V FS    P++GNTEA +M+K+D
Sbjct: 180 LLQEAGAKRLIPLKVSGPFHTALLEPASQKLAETLAQVSFSDFTCPLVGNTEAAVMQKED 239

Query: 241 IKSLLARQVMEPVRFDESIETMKKMGMTQVVEIGPGKVLSGFLKKIDSSLSVHSVEDKIG 300
           I  LL RQV EPVRF ESI  M++ G++  +EIGPGKVLSGF+KKID +  + VED+
Sbjct: 240 IAQLLTRQVKEPVRFYESIGVMQEAGISNFIEIGPGKVLSGFVKKIDQTAHLAHVEDQAS 299

Query: 301 FNNLKE 306
              L E
Sbjct: 300 LVALLE 305
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5195> which encodes the amino acid sequence <SEQ ID 5196>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1602(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/299 (67%), Positives = 248/299 (82%), Gaps = 1/299 (0%)

Query: 1    MNKVSFLFAGQGAQKLGMARDLYETFPIVKETFDKASHVLGYDLRELIDKDLDKLNQTKY 60
            M K +FLFAGQGAQKLGMARD Y+ F IV++TFD+AS VLGYDLR LID D  KLNQT Y
Sbjct: 3    MTKTAFLFAGQGAQKLGMARDFYDNFAIVRKTFDQASQVLGYDLRRLIDSDELKLNQTSY 62

Query: 61   TQPAILTTSTAIYRLILKEIELRPDMVAGLSLGEYSALVASGAIRFEDAVVLVARRGQLM 120
            TQPAILT+S AIYR +L    ++PDMVAGLSLGEYSALVASGA+ FED + LVA+RG+LM
Sbjct: 63   TQPAILTSSIAIYR-VLGLHHVKPDMVAGLSLGEYSALVASGALSFEDTLSLVAKRGRLM 121

Query: 121 EAAAPAGSGKMVAVLNADRQIIEDACKKASQFGIVSPANYNTPKQIVIGGESIAVNAAVE 180
            E AAP GSGKMVAV+N D Q+IE+ C+ A++ G+V+PANYNTP QIVIGG++ AVN AVE
Sbjct: 122 EEAAPQGSGKMVAVMNTDVQVIEEVCQIAAKHGVVAPANYNTPSQIVIGGQTDAVNVAVE 181

Query: 181 ELKQQGVKRLIPLNVSGPFHTALLKPASQKLSDVLDKVHFSVSEIPVIGNTEAQIMKKDD 240
             LK++GVKRLIPLNVSGPFHTALL+PAS+ L+  L++ +FS  +IP++GNTEA IM+KD
Sbjct: 182 LLKERGVKRLIPLNVSGPFHTALLEPASRLLAKELERYNFSDFKIPLVGNTEANIMEKDR 241

Query: 241 IKSLLARQVMEPVRFDESIETMKKMGMTQVVEIGPGKVLSGFLKKIDSSLSVHSVEDKI 299
            I  LLARQVMEPVRF +S+ T+ + G+TQ +E+GPGKVL+GF+KKID +L   SVE+ +
Sbjct: 242 IPELLARQVMEPVRFYDSVATLVESGITQFIEVGPGKVLTGFVKKIDKNLLCTSVENMV 300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1675

A DNA sequence (GBSx1779) was identified in *S.agalactiae* <SEQ ID 5197> which encodes the amino acid sequence <SEQ ID 5198>. This protein is predicted to be beta-ketoacyl-ACP reductase (fabG). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0930(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98275 GB:AF197933 beta-ketoacyl-ACP reductase [Streptococcus pneumoniae]
 Identities = 184/243 (75%), Positives = 212/243 (86%)

Query: 1   MQLKDKNIFITGSSRGIGLAIAHQFAQLGANIVLNGRSEISEDLIAEFADYGVKVIAISG 60
            M+L+ KNIFITGSSRGIGLAIAH+FAQ GANIVLN R  ISE+L+AEF++YG+KV+ ISG
Sbjct: 1   MKLEHKNIFITGSSRGIGLAIAHKFAQAGANIVLNSRGAISEELLAEFSNYGIKVVPISG 60

Query: 61  DVSSFEDANRMIKEAIASLGSVDVLVNNAGITNDKLMLKMTVEDFESVLKINLTGAFNMT 120
            DVS F DA RMI +AIA LGSVDVLVNNAGIT D LMLKMT  DFE VLK+NLTGAFNMT
Sbjct: 61  DVSDFADAKRMIDQAIAELGSVDVLVNNAGITQDTLMLKMTEADFEKVLKVNLTGAFNMT 120

Query: 121 QSVLKPMTKARQGAIINISSVVGLTGNVGQANYAASKAGLIGFTKSVAREVAARGIRVNA 180
            QSVLKPM KAR+GAIIN+SSVVGL GN+GQANYAASKAGLIGFTKSVAREVA+R IRVN
Sbjct: 121 QSVLKPMMKAREGAIINMSSVVGLMGNIGQANYAASKAGLIGFTKSVAREVASRNIRVNV 180

Query: 181 IAPGFIESDMTDVIPEKMQEAILAQIPMKRIGKGKEVAQVASFLAEQEYLTGQVIAIDGG 240
            IAPG IESDMT ++ +K++EA LAQIPMK  G+ ++VA +  FLA Q+YLTGQV+AIDGG
Sbjct: 181 IAPGMIESDMTAILSDKIKEATLAQIPMKEFGQAEQVADLTVFLAGQDYLTGQVVAIDGG 240

Query: 241 MTM 243
            ++M
Sbjct: 241 LSM 243
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3865> which encodes the amino acid sequence <SEQ ID 3866>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/244 (82%), Positives = 220/244 (89%)

Query: 1   MQLKDKNIFITGSSRGIGLAIAHQFAQLGANIVLNGRSEISEDLIAEFADYGVKVIAISG 60
            M++K KNIFITGS+RGIGLA+AHQFA L ANIVLNGRS ISE+L+A F DYGV V+ ISG
Sbjct: 1   MEIKGKNIFITGSTRGIGLAMAHQFASLEANIVLNGRSAISEELVASFTDYGVTVVTISG 60

Query: 61  DVSSFEDANRMIKEAIASLGSVDVLVNNAGITNDKLMLKMTVEDFESVLKINLTGAFNMT 120
            DVS   +A RM+ EAI SLGS+DVLVNNAGITNDKLMLKMT EDFE VLKINLTGAFNMT
Sbjct: 61  DVSEASEAKRMVNEAIESLGSIDVLVNNAGITNDKLMLKMTEEDFERVLKINLTGAFNMT 120

Query: 121 QSVLKPMTKARQGAIINISSVVGLTGNVGQANYAASKAGLIGFTKSVAREVAARGIRVNA 180
            QSVLKPM KARQGAIIN+SSVVGLTGN+GQANYAASKAG+IGFTKSVAREVAAR I VNA
Sbjct: 121 QSVLKPMIKARQGAIINVSSVVGLTGNIGQANYAASKAGMIGFTKSVAREVAARNICVNA 180

Query: 181 IAPGFIESDMTDVIPEKMQEAILAQIPMKRIGKGKEVAQVASFLAEQEYLTGQVIAIDGG 240
            IAPGFIESDMT V+PEKMQE IL+QIPMKRIGK +EVA +ASFL EQ+Y+TGQVIAIDGG
Sbjct: 181 IAPGFIESDMTGVLPEKMQEQILSQIPMKRIGKAQEVAHLASFLVEQDYITGQVIAIDGG 240

Query: 241 MTMQ 244
            MTMQ
Sbjct: 241 MTMQ 244
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1676

A DNA sequence (GBSx1780) was identified in *S.agalactiae* <SEQ ID 5199> which encodes the amino acid sequence <SEQ ID 5200>. This protein is predicted to be 3-oxoacyl-(acyl-carrier-protein) synthase II (fabF). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  338 - 354 ( 338 - 354)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98276 GB:AF197933 beta-ketoacyl-ACP synthase II
           [Streptococcus pneumoniae]
 Identities = 340/410 (82%), Positives = 375/410 (90%)

Query: 1   MTLQRVVVTGYGVTSPIGNTPEEFWNSLKEGNVGIGPITKFDSSDFMVKNAAEIHDFPFD 60
           M L RVVVTGYGVTSPIGNTPEEFWNSL  G +GIG ITKFD SDF V NAAEI DFPFD
Sbjct: 1   MKLNRVVVTGYGVTSPIGNTPEEFWNSLATGKIGIGGITKFDHSDFDVHNAAEIQDFPFD 60

Query: 61  KYFVKKDLNRFDMYSLYALYASSEAIQHANLNLDEIDADRFGVIVASGIGGIQEIEEQVI 120
           KYFVKKD NRFD YSLYALYA+ EA+ HANL+++ ++ DRFGVIVASGIGGI+EIE+QV+
Sbjct: 61  KYFVKKDTNRFDNYSLYALYAAQEAVNHANLDVEALNRDRFGVIVASGIGGIKEIEDQVL 120

Query: 121 RLHEKGPKRVKPMTLPKALPNMAAGNVAMRLGAHGVCKSINTACASSNDAIGDAFRNIKF 180
           RLHEKGPKRVKPMTLPKALPNMA+GNVAMR GA+GVCKSINTAC+SSNDAIGDAFR+IKF
Sbjct: 121 RLHEKGPKRVKPMTLPKALPNMASGNVAMRFGANGVCKSINTACSSSNDAIGDAFRSIKF 180

Query: 181 GIQDIMVVGGAEAAITKFAIAGFQSLTALSTTEDPSRASIPFDKDRNGFIMGEGSGMLVL 240
           G QD+M+VGG EA+IT FAIAGFQ+LTALSTTEDP+RASIPFDKDRNGF+MGEGSGMLVL
Sbjct: 181 GFQDVMLVGGTEASITPFAIAGFQALTALSTTEDPTRASIPFDKDRNGFVMGEGSGMLVL 240

Query: 241 ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGLGATKAIQLALVEANIKPEEVNYV 300
           ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEG GA KAI+LAL EA I PE+V YV
Sbjct: 241 ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGQGAIKAIKLALEEAEISPEQVAYV 300

Query: 301 NAHGTSTPANEKGESQAIVAALGTDVPVSSTKSFTGHLLGAAGAVEAIATIEAIRHSYVP 360
           NAHGTSTPANEKGES AIVA LG +VPVSSTKSFTGHLLGAAGAVEAI TIEA+RH++VP
Sbjct: 301 NAHGTSTPANEKGESGAIVAVLGKEVPVSSTKSFTGHLLGAAGAVEAIVTIEAMRHNFVP 360

Query: 361 MTAGTTELSEDITANVIFGQGQDADIRYAISNTFGFGGHNAVLAFKRWED 410
           MTAGT+E+S+ I ANV++GQG + +I YAISNTFGFGGHNAVLAFKRWE+
Sbjct: 361 MTAGTSEVSDYIEANVVYGQGLEKEIPYAISNTFGFGGHNAVLAFKRWEN 410
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3851> which encodes the amino acid sequence <SEQ ID 3852>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0890(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 346/410 (84%), Positives = 377/410 (91%)

Query: 1   MTLQRVVVTGYGVTSPIGNTPEEFWNSLKEGNVGIGPITKFDSSDFMVKNAAEIHDFPFD 60
           MT +RVVVTGYG+TSPIG+ PE FWN+LK G +GIGPITKFD++D+ VKNAAEI DFPFD
Sbjct: 1   MTFKRVVVTGYGLTSPIGHDPETFWNNLKAGQIGIGPITKFDTTDYAVKNAAEIQDFPFD 60

Query: 61  KYFVKKDLNRFDMYSLYALYASSEAIQHANLNLDEIDADRFGVIVASGIGGIQEIEEQVI 120
           KYFVKKDLNRFD YSLYALYA+ EAI HA+LN++ +D+DRFGVIVASGIGGI EIEEQVI
Sbjct: 61  KYFVKKDLNRFDRYSLYALYAAKEAINHADLNIEMVDSDRFGVIVASGIGGIAEIEEQVI 120

Query: 121 RLHEKGPKRVKPMTLPKALPNMAAGNVAMRLGAHGVCKSINTACASSNDAIGDAFRNIKF 180
           RLHEKGPKRVKPMTLPKALPNMAAGNVAM L A GVCKSINTACASSNDAIGDAFR IKF
Sbjct: 121 RLHEKGPKRVKPMTLPKALPNMAAGNVAMSLKAQGVCKSINTACASSNDAIGDAFRAIKF 180

Query: 181 GIQDIMVVGGAEAAITKFAIAGFQSLTALSTTEDPSRASIPFDKDRNGFIMGEGSGMLVL 240
           G QD+M+VGG+EAAITKFAIAGFQSLTALSTTEDPSR+SIPFDKDRNGFIMGEGSGMLVL
Sbjct: 181 GTQDVMIVGGSEAAITKFAIAGFQSLTALSTTEDPSRSSIPFDKDRNGFIMGEGSGMLVL 240

Query: 241 ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGLGATKAIQLALVEANIKPEEVNYV 300
           ESLEHA++RGATILAE+VGYGNTCDAYHMTSP+PEGLGA KAI LAL EA I+   +NYV
Sbjct: 241 ESLEHAQERGATILAEIVGYGNTCDAYHMTSPNPEGLGARKAIHLALQEAGIEASAINYV 300

Query: 301 NAHGTSTPANEKGESQAIVAALGTDVPVSSTKSFTGHLLGAAGAVEAIATIEAIRHSYVP 360
           NAHGTSTPANEKGESQAIVA LG DVPVSSTKSFTGHLLGAAGA+EAIATIEA+RH+YVP
Sbjct: 301 NAHGTSTPANEKGESQAIVAVLGKDVPVSSTKSFTGHLLGAAGAIEAIATIEAMRHNYVP 360

Query: 361 MTAGTTELSEDITANVIFGQGQDADIRYAISNTFGFGGHNAVLAFKRWED 410
           MTAGT  LSEDI ANVIFG+G++  I YAISNTFGFGGHNAVLAFK WE+
Sbjct: 361 MTAGTQALSEDIEANVIFGEGKETAINYAISNTFGFGGHNAVLAFKCWEE 410
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1677

A DNA sequence (GBSx1781) was identified in *S.agalactiae* <SEQ ID 5201> which encodes the amino acid sequence <SEQ ID 5202>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3052(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9553> which encodes amino acid sequence <SEQ ID 9554> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98277 GB:AF197933 biotin carboxyl carrier protein
           [Streptococcus pneumoniae]
 Identities = 103/169 (60%), Positives = 127/169 (74%), Gaps = 11/169 (6%)

Query: 19  LDIQEIKDLMTQFDESSLREFSFKTSDGELSFSKNEGKAPLVPTMSPMSHQPEATPTIAT 78
           +++ +IKDLMTQFD+SSLREFS+K    EL FSKNE +   VP ++    Q    P +AT
Sbjct: 1   MNLNDIKDLMTQFDQSSLREFSYKNGTDELQFSKNEARP--VPEVAT---QVAPAPVLAT 55

Query: 79  PVSNEAGEQTKQATEVVSEIP---ESTVTVAEGDVVESPLVGVAYLASGPDKPNFVSVGD 135
           P  +     +  A  V  E+P   E++V   EG++VESPLVGV YLA+GPDKP FV+VGD
Sbjct: 56  P--SPVAPTSAPAETVAEEVPAPAEASVAT-EGNLVESPLVGVVYLAAGPDKPAFVTVGD 112
```

```
Query: 136 SVKKGQTLMIIEAMKVMNEVPAPHDGVVTEILVANEEVIEFGKGLVRIK 184
           SVKKGQTL+IIEAMKVMNE+PAP DGVVTEILV+NEE++EFGKGLVRIK
Sbjct: 113 SVKKGQTLVIIEAMKVMNEIPAPKDGVVTEILVSNEEMVEFGKGLVRIK 161
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5203> which encodes the amino acid sequence <SEQ ID 5204>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3132(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 107/171 (62%), Positives = 126/171 (73%), Gaps = 10/171 (5%)

Query: 19  LDIQEIKDLMTQFDESSLREFSFKTSDGELSFSKNEGKAPLVPTMSPMSHQPEATPT--- 75
           L+IQEIKDLM QFD SSLREF FKT++GEL FSKNE         +  S+Q  A P
Sbjct: 1   LNIQEIKDLMAQFDTSSLREFLFKTNEGELIFSKNEQHLN-----ASTSNQEHAVPVPQV 55

Query: 76  --IATPVSNEAGEQTKQATEVVSEIPESTVTVAEGDVVESPLVGVAYLASGPDKPNFVSV 133
             +  P ++EA          V E P++   VAEGD+VESPLVGVAYLA+ PDKP FV+V
Sbjct: 56  QLVPNPTASEASSPASVKDVPVEEQPQAESFVAEGDIVESPLVGVAYLAASPDKPPFVAV 115

Query: 134 GDSVKKGQTLMIIEAMKVMNEVPAPHDGVVTEILVANEEVIEFGKGLVRIK 184
           GD+VKKGQTL+IIEAMKVMNEVPAP DGV+TEILV+NE+VIEFG+GLVRIK
Sbjct: 116 GDTVKKGQTLVIIEAMKVMNEVPAPCDGVITEILVSNEDVIEFGQGLVRIK 166
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1678

A DNA sequence (GBSx1782) was identified in *S.agalactiae* <SEQ ID 5205> which encodes the amino acid sequence <SEQ ID 5206>. This protein is predicted to be beta-hydroxyacyl-ACP dehydratase (fabZ). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2267(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98278 GB:AF197933 beta-hydroxyacyl-ACP dehydratase
          [Streptococcus pneumoniae]
 Identities = 130/140 (92%), Positives = 135/140 (95%)

Query: 1   MIDIKEIREALPHRYPMLLVDRVLEVSEDEIVAIKNVSINEPFFNGHFPEYPVMPGVLIM 60
           MIDI+ I+EALPHRYPMLLVDRVLEVSED IVAIKNV+INEPFFNGHFP+YPVMPGV+IM
Sbjct: 1   MIDIQGIKEALPHRYPMLLVDRVLEVSEDTIVAIKNVTINEPFFNGHFPQYPVMPGVVIM 60

Query: 61  EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPGDQLVMTAKFVKRRGTIAVVEA 120
           EALAQTAGVLELSK ENKGKLVFYAGMDKVKFKKQVVPGDQLVMTA FVKRRGTIAVVEA
Sbjct: 61  EALAQTAGVLELSKPENKGKLVFYAGMDKVKFKKQVVPGDQLVMTATFVKRRGTIAVVEA 120

Query: 121 IAEVDGKLAASGTLTFAIGN 140
            AEVDGKLAASGTLTFAIGN
```

```
Sbjct: 121 KAEVDGKLAASGTLTFAIGN 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5207> which encodes the amino acid sequence <SEQ ID 5208>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1882(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 127/139 (91%), Positives = 133/139 (95%)

Query: 1   MIDIKEIREALPHRYPMLLVDRVLEVSEDEIVAIKNVSINEPFFNGHFPEYPVMPGVLIM 60
            M+DI+EI+ ALPHRYPMLLVDRVLEVS+D IVAIKNV+INEPFFNGHFP YPVMPGVLIM
Sbjct: 1   MMDIREIQAALPHRYPMLLVDRVLEVSDDHIVAIKNVTINEPFFNGHFPHYPVMPGVLIM 60

Query: 61  EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPGDQLVMTAKFVKRRGTIAVVEA 120
            EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPGDQLVMTA F+KRRGTIAVVEA
Sbjct: 61  EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPGDQLVMTATFIKRRGTIAVVEA 120

Query: 121 IAEVDGKLAASGTLTFAIG 139
             AEVDGKLAASGTLTFA G
Sbjct: 121 RAEVDGKLAASGTLTFACG 139
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1679

A DNA sequence (GBSx1783) was identified in *S.agalactiae* <SEQ ID 5209> which encodes the amino acid sequence <SEQ ID 5210>. This protein is predicted to be acetyl-coenzyme A carboxylase, biotin carboxylase (accC). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1203(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98279 GB:AF197933 acetyl-CoA carboxylase biotin carboxylase
          subunit [Streptococcus pneumoniae]
 Identities = 361/451 (80%), Positives = 405/451 (89%)

Query: 1   MFKKILIANRGEIAVRIIRAAREMGISTVAIYSEADKESLHTILADEAICVGPAKSAESY 60
            MF+KILIANRGEIAVRIIRAARE+GI+TVA+YS ADKE+LHT+LADEA+C+GP K+ ESY
Sbjct: 1   MFRKILIANRGEIAVRIIRAARELGIATVAVYSTADKEALHTLLADEAVCIGPGKATESY 60

Query: 61  LNVNAILSAAIVTGAEAVHPGFGFLSENSKFATMCEEMNLKFIGPSGEVMDKMGDKINAR 120
            LN+NA+LSAA++T AEA+HPGFGFLSENSKFATMCEE+ +KFIGPSG VMD MGDKINAR
Sbjct: 61  LNINAVLSAAVLTEAEAIHPGFGFLSENSKFATMCEEVGIKFIGPSGHVMDMMGDKINAR 120

Query: 121 TEMIKADVPVIPGSDGQVTSVEEAVSIAEEIGYPLMLKASAGGGGKGIRKVKSADELKPA 180
             +MIKA VPVIPGSDG+V + EEA+ +AE+IGYP+MLKASAGGGGKGIRKV+  D+L  A
Sbjct: 121 AQMIKAGVPVIPGSDGEVHNSEEALIVAEKIGYPVMLKASAGGGGKGIRKVEKPDDLVSA 180
```

```
Query: 181 FESASQEALAAFGNGAMYIEKVIYPARHIEVQILGDSFGKIVHLGERDCSLQRNNQKVLE 240
           FE+AS EA A +GNGAMYIE+VIYPARHIEVQILGD  G ++HLGERDCSLQRNNQKVLE
Sbjct: 181 FETASSEAKANYGNGAMYIERVIYPARHIEVQILGDEHGHVIHLGERDCSLQRNNQKVLE 240

Query: 241 ESPSVAIGNTLRQQIGEAAVRAAEAVSYENAGTIEFLLDENSGQFYFMEMNTRVQVEHPV 300
           ESPS+AIG TLR +IG AAVRAAE V YENAGTIEFLLDE S  FYFMEMNTRVQVEHPV
Sbjct: 241 ESPSIAIGKTLRHEIGAAAVRAAEFVGYENAGTIEFLLDEASSNFYFMEMNTRVQVEHPV 300

Query: 301 TEFVTGVDIVKEQIRIAAGIPLSVSQNDIKLTGHAIECRINAENPQFNFAPCPGTINGLH 360
           TEFV+GVDIVKEQI IAAG PLSV Q DI L GHAIECRINAENP FNFAP PG I  L+
Sbjct: 301 TEFVSGVDIVKEQICIAAGQPLSVKQEDIVLRGHAIECRINAENPAFNFAPSPGKITNLY 360

Query: 361 LPAGGMGLRVDSAVYTGYTIPPYYDSMIAKVIVHGENRFDALMKMQRALYELEIDGIVTN 420
           LP+GG+GLRVDSAVY GYTIPPYYDSMIAK+IVHGENRFDALMKMQRALYELEI+G+ TN
Sbjct: 361 LPSGGVGLRVDSAVYPGYTIPPYYDSMIAKIIVHGENRFDALMKMQRALYELEIEGVQTN 420

Query: 421 TEFQMDLISDKKVLAGDYDTSFLMEDFLPRY 451
            +FQ+DLISD+ V+AGDYDTSFLME FLP+Y
Sbjct: 421 ADFQLDLISDRNVIAGDYDTSFLMETFLPKY 451
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5211> which encodes the amino acid sequence <SEQ ID 5212>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1784(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 369/451 (81%), Positives = 421/451 (92%)

Query: 1   MFKKILIANRGEIAVRIIRAAREMGISTVAIYSEADKESLHTILADEAICVGPAKSAESY 60
           MFKKILIANRGEIAVRIIRAARE+GISTVA+YSEADKE+LHTILADEAIC+GPA+S ESY
Sbjct: 17  MFKKILIANRGEIAVRIIRAARELGISTVAVYSEADKEALHTILADEAICIGPARSKESY 76

Query: 61  LNVNAILSAAIVTGAEAVHPGFGFLSENSKFATMCEEMNLKFIGPSGEVMDKMGDKINAR 120
           LN+N++LSAAIVTGA+A+HPGFGFLSENSKFATMCEEMN+KFIGPS  VMDKMGDKINAR
Sbjct: 77  LNMNSVLSAAIVTGAQAIHPGFGFLSENSKFATMCEEMNIKFIGPSASVMDKMGDKINAR 136

Query: 121 TEMIKADVPVIPGSDGQVTSVEEAVSIAEEIGYPLMLKASAGGGGKGIRKVKSADELKPA 180
           +EMIKA VPVIPGSDG+V + +EA++IA +IGYP+MLKASAGGGGKGIRKV++  +L+ A
Sbjct: 137 SEMIKAGVPVIPGSDGEVYNAQEALAIANKIGYPVMLKASAGGGGKGIRKVETEADLEAA 196

Query: 181 FESASQEALAAFGNGAMYIEKVIYPARHIEVQILGDSFGKIVHLGERDCSLQRNNQKVLE 240
           F +ASQEAL AFGNGAMY+EKVIYPARHIEVQILGD++G I+HLGERDCSLQRNNQKVLE
Sbjct: 197 FNAASQEALGAFGNGAMYLEKVIYPARHIEVQILGDAYGNIIHLGERDCSLQRNNQKVLE 256

Query: 241 ESPSVAIGNTLRQQIGEAAVRAAEAVSYENAGTIEFLLDENSGQFYFMEMNTRVQVEHPV 300
           ESPS+AIGNTLR ++G+AAVRAAEAV+YENAGTIEFLLDE+S +FYFMEMNTR+QVEHPV
Sbjct: 257 ESPSIAIGNTLRHEMGQAAVRAAEAVAYENAGTIEFLLDEDSEKFYFMEMNTRIQVEHPV 316

Query: 301 TEFVTGVDIVKEQIRIAAGIPLSVSQNDIKLTGHAIECRINAENPQFNFAPCPGTINGLH 360
           TEFVTGVDIVKEQI+IAAG PL+++Q DI +TGHAIECRINAEN  FNFAP PG I  L+
Sbjct: 317 TEFVTGVDIVKEQIKIAAGQPLAINQEDITITGHAIECRINAENTAFNFAPSPGKITDLY 376

Query: 361 LPAGGMGLRVDSAVYTGYTIPPYYDSMIAKVIVHGENRFDALMKMQRALYELEIDGIVTN 420
           +P+GG+GLRVDSAVY GY IPPYYDSMIAK+IVHG NRFDALMKMQRAL ELEI+GI+TN
Sbjct: 377 MPSGGVGLRVDSAVYNGYAIPPYYDSMIAKIIVHGSNRFDALMKMQRALVELEIEGIITN 436

Query: 421 TEFQMDLISDKKVLAGDYDTSFLMEDFLPRY 451
           T+FQ+DLISDK+V+AGDYDTSFLME FLP Y
Sbjct: 437 TDFQLDLISDKRVIAGDYDTSFLMETFLPHY 467
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1680

A DNA sequence (GBSx1784) was identified in *S.agalactiae* <SEQ ID 5213> which encodes the amino acid sequence <SEQ ID 5214>. This protein is predicted to be acetyl-CoA carboxylase beta subunit (accD). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3571(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98280 GB:AF197933 acetyl-CoA carboxylase beta subunit
          [Streptococcus pneumoniae]
 Identities = 221/285 (77%), Positives = 248/285 (86%), Gaps = 1/285 (0%)

Query: 1   MALFSKKDKYIRISPNKALGSSDKRSLPEVPDELFAKCPSCKHMIYQKDLGLAKICPACS 60
           MALFSKKDKYIRI+PN+++    +   PEVPDELF++CP CKH IYQKDLG  +ICP CS
Sbjct: 1   MALFSKKDKYIRINPNRSVREKPQAK-PEVPDELFSQCPGCKHTIYQKDLGSERICPHCS 59

Query: 61  YNFRISAQERLLLTVDEDSFEELFTGIETKDPLNFPNYREKLAATRQKTNLDEAVVTGLA 120
           Y FRISAQERL LT+D  +F+ELFTGIE+KDPL+FP Y++KLA+ R+KT L EAVVTG A
Sbjct: 60  YTFRISAQERLALTIDMGTFKELFTGIESKDPLHFPGYQKKLASMREKTGLHEAVVTGTA 119

Query: 121 KIKGQTTALAIMDSHFIMASMGTVVGEKLTRLFELATEKKLPIVIFTASGGARMQEGIMS 180
            IKGQT AL IMDS+FIMASMGTVVGEK+TRLFE AT +KLP+V+FTASGGARMQEGIMS
Sbjct: 120 LIKGQTVALGIMDSNFIMASMGTVVGEKITRLFEYATVEKLPVVLFTASGGARMQEGIMS 179

Query: 181 LMQMAKVSAAVKRHSNQGLFYLTILTDPTTGGVTASFAMEGDIILAEPQALVGFAGRRVI 240
           LMQMAK+SAAVKRHSN GLFYLTILTDPTTGGVTASFAMEGDIILAEPQ+LVGFAGRRVI
Sbjct: 180 LMQMAKISAAVKRHSNAGLFYLTILTDPTTGGVTASFAMEGDIILAEPQSLVGFAGRRVI 239

Query: 241 ETTVREDLPEGFQKAEFLLEHGFVDAIINRTELRDCIAQLIAFHG 285
           E TVRE LPE FQKAEFLLEHGFVDAI+ R +L D IA L+  HG
Sbjct: 240 ENTVRESLPEDFQKAEFLLEHGFVDAIVKRRDLPDTIASLVRLHG 284
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5215> which encodes the amino acid sequence <SEQ ID 5216>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4092(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 232/285 (81%), Positives = 253/285 (88%)

Query: 1   MALFSKKDKYIRISPNKALGSSDKRSLPEVPDELFAKCPSCKHMIYQKDLGLAKICPACS 60
           MALF KKDKYIRI+PN +L  S   ++PEVPDELFAKCP+CKHMIY+KDLGLAKICP CS
Sbjct: 1   MALFRKKDKYIRITPNNSLKGSVSHNVPEVPDELFAKCPACKHMIYKKDLGLAKICPTCS 60

Query: 61  YNFRISAQERLLLTVDEDSFEELFTGIETKDPLNFPNYREKLAATRQKTNLDEAVVTGLA 120
           YNFRISAQERL LTVDE SF+ELFT IETKDPL FP Y+EKL   ++ T L EAV+TG A
```

```
Sbjct: 61  YNFRISAQERLTLTVDEGSFQELFTSIETKDPLRFPGYQEKLQKAKETTGLHEAVLTGKA 120

Query: 121 KIKGQTTALAIMDSHFIMASMGTVVGEKLTRLFELATEKKLPIVIFTASGGARMQEGIMS 180
             +K Q  ALAIMDSHFIMASMGTVVGEK+TRLFELA E+ LP+VIFTASGGARMQEGIMS
Sbjct: 121 MVKEQKIALAIMDSHFIMASMGTVVGEKITRLFELAIEENLPVVIFTASGGARMQEGIMS 180

Query: 181 LMQMAKVSAAVKRHSNQGLFYLTILTDPTTGGVTASFAMEGDIILAEPQALVGFAGRRVI 240
            LMQMAKVSAAVKRHSN GLFYLTILTDPTTGGVTASFAMEGDIILAEPQ+LVGFAGRRVI
Sbjct: 181 LMQMAKVSAAVKRHSNAGLFYLTILTDPTTGGVTASFAMEGDIILAEPQSLVGFAGRRVI 240

Query: 241 ETTVREDLPEGFQKAEFLLEHGFVDAIINRTELRDCIAQLIAFHG 285
            ETTVRE+LP+ FQKAEFL +HGFVDAI+ RTELRD IA L+AFHG
Sbjct: 241 ETTVRENLPDDFQKAEFLQDHGFVDAIVKRTELRDKIAHLVAFHG 285
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1681

A DNA sequence (GBSx1785) was identified in *S.agalactiae* <SEQ ID 5217> which encodes the amino acid sequence <SEQ ID 5218>. This protein is predicted to be acetyl-CoA carboxylase alpha subunit (accA). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane  149 - 165 ( 149 - 165)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9555> which encodes amino acid sequence <SEQ ID 9556> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98281 GB:AF197933 acetyl-CoA carboxylase alpha subunit
          [Streptococcus pneumoniae]
 Identities = 186/254 (73%), Positives = 222/254 (87%)

Query: 13  DVTRILKDARDQGRLTALDYAELIFDNFMELHGDRQFADDKSIIGGLGYLAGRPVTIVGI 72
            ++ +I+++AR+Q RLT LD+A  IFD F++LHGDR F DD +++GG+G+L  + VT+VGI
Sbjct: 2   NIAKIVREAREQSRLTTLDFATGIFDEFIQLHGDRSFRDDGAVVGGIGWLGDQAVTVVGI 61

Query: 73  QKGKNLQDNLDRHFGQPHPEGYRKALRLMKQAEKFGRPVITFINTAGAYPGVGAEERGQG 132
            QKGK+LQDNL R+FGQPHPEGYRKALRLMKQAEKFGRPV+TFINTAGAYPGVGAEERGQG
Sbjct: 62  QKGKSLQDNLKRNFGQPHPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 121

Query: 133 EAIARNLLEMSDLKVPIIAIIIGEGGSGGALALAVADKVWMLEHTVYSILSPEGFASILW 192
            EAIARNL+EMSDLKVPIIAIIIGEGGSGGALALAVAD+VWMLE+++Y+ILSPEGFASILW
Sbjct: 122 EAIARNLMEMSDLKVPIIAIIIGEGGSGGALALAVADRVWMLENSIYAILSPEGFASILW 181

Query: 193 KDGTRTTEAAQLMKMTAGELYHMEVVDKVIPEHGYFSSEIVDMIKTSLISELEVLSQLSL 252
            KDGTR  EAA+LMK+T+ EL  M+VVDKVI E G  S E++  +K  L +EL  LSQ  L
Sbjct: 182 KDGTRAMEAAELMKITSHELLEMDVVDKVISEIGLSSKELIKSVKKELQTELARLSQKPL 241

Query: 253 EDLLEQRYQRFRKY 266
            E+LLE+RYQRFRKY
Sbjct: 242 EELLEERYQRFRKY 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5219> which encodes the amino acid sequence <SEQ ID 5220>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane  139 - 155 ( 139 - 155)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF98281 GB:AF197933 acetyl-CoA carboxylase alpha subunit
           [Streptococcus pneumoniae]
 Identities = 189/254 (74%), Positives = 225/254 (88%)

Query: 3   DVSRILKEARDQGRLTTLDYANLIFDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTVIGI 62
           ++++I++EAR+Q RLTTLD+A  IFD+F++LHGDR F DDGA+VGG+ +L  Q VTV+GI
Sbjct: 2   NIAKIVREAREQSRLTTLDFATGIFDEFIQLHGDRSFRDDGAVVGGIGWLGDQAVTVVGI 61

Query: 63  QKGKNLQDNLARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 122
           QKGK+LQDNL RNFGQP+PEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG
Sbjct: 62  QKGKSLQDNLKRNFGQPHPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 121

Query: 123 EAIAKNLMEMSDLKVPIIAIIIGEGGSGGALALAVADQVWMLENTMYAVLSPEGFASILW 182
           EAIA+NLMEMSDLKVPIIAIIIGEGGSGGALALAVAD+VWMLEN++YA+LSPEGFASILW
Sbjct: 122 EAIARNLMEMSDLKVPIIAIIIGEGGSGGALALAVADRVWMLENSIYAILSPEGFASILW 181

Query: 183 KDGSRATEAAELMKITAGELYKMGIVDRIIPEHGYFSSEIVDIIKANLIEQITSLQAKPL 242
           KDG+RA EAAELMKIT+ EL +M +VD++I E G  S E++  +K  L  ++  L  KPL
Sbjct: 182 KDGTRAMEAAELMKITSHELLEMDVVDKVISEIGLSSKELIKSVKKELQTELARLSQKPL 241

Query: 243 DQLLDERYQRFRKY 256
           ++LL+ERYQRFRKY
Sbjct: 242 EELLEERYQRFRKY 255
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 204/254 (80%), Positives = 236/254 (92%)

Query: 13  DVTRILKDARDQGRLTALDYAELIFDNFMELHGDRQFADDKSIIGGLGYLAGRPVTIVGI 72
           DV+RILK+ARDQGRLT LDYA LIFD+FMELHGDR F+DD +I+GGL YLAG+PVT++GI
Sbjct: 3   DVSRILKEARDQGRLTTLDYANLIFDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTVIGI 62

Query: 73  QKGKNLQDNLDRHFGQPHPEGYRKALRLMKQAEKFGRPVITFINTAGAYPGVGAEERGQG 132
           QKGKNLQDNL R+FGQP+PEGYRKALRLMKQAEKFGRPV+TFINTAGAYPGVGAEERGQG
Sbjct: 63  QKGKNLQDNLARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 122

Query: 133 EAIARNLLEMSDLKVPIIAIIIGEGGSGGALALAVADKVWMLEHTVYSILSPEGFASILW 192
           EAIA+NL+EMSDLKVPIIAIIIGEGGSGGALALAVAD+VWMLE+T+Y++LSPEGFASILW
Sbjct: 123 EAIAKNLMEMSDLKVPIIAIIIGEGGSGGALALAVADQVWMLENTMYAVLSPEGFASILW 182

Query: 193 KDGTRTTEAAQLMKMTAGELYHMEVVDKVIPEHGYFSSEIVDMIKTSLISELEVLSQLSL 252
           KDG+R TEAA+LMK+TAGELY M +VD++IPEHGYFSSEIVD+IK +LI ++   L    L
Sbjct: 183 KDGSRATEAAELMKITAGELYKMGIVDRIIPEHGYFSSEIVDIIKANLIEQITSLQAKPL 242

Query: 253 EDLLEQRYQRFRKY 266
           + LL++RYQRFRKY
Sbjct: 243 DQLLDERYQRFRKY 256
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1682

A DNA sequence (GBSx1786) was identified in *S.agalactiae* <SEQ ID 5221> which encodes the amino acid sequence <SEQ ID 5222>. This protein is predicted to be sakacin A production response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3304(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9557> which encodes amino acid sequence <SEQ ID 9558> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA88824 GB:AB016077 sakacin A production response regulator
          [Streptococcus mutans]
 Identities = 76/142 (53%), Positives = 99/142 (69%)

Query: 36  MQTFKAKGQLARNSFTELSRALEQRMDGFKMQRVSNWANQAQVGRPHFWVYYRKDTDQLD 95
           M   K  GQ AR  FTE+++ L  ++  F+M RVSNWANQAQV RPHFW YY++  D  D
Sbjct: 1   MIALKTLGQSARAEFTEIAKVLALKVSPFEMMRVSNWANQAQVVRPHFWCYYKQPEDNQD 60

Query: 96  DVAVALRVYGVKDSFGVSLEVSFVERQKSDKTLEKQARVLSIPIASPLYFMVQRQGETHR 155
           DV +A+R+YG   +FG+S+EVSF+ER+KS  TL KQ +VL IPIA PLY+  Q + E+HR
Sbjct: 61  DVGLAIRLYGNSANFGISVEVSFIERKKSKATLAKQHKVLDIPIAEPLYYFAQEKSESHR 120

Query: 156 EEGNEENRQRLMQEIKSGKVRK 177
             G E  RQ L Q++  G+VRK
Sbjct: 121 VSGTEAYRQMLRQKVADGQVRK 142
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1683

A DNA sequence (GBSx1787) was identified in *S.agalactiae* <SEQ ID 5223> which encodes the amino acid sequence <SEQ ID 5224>. This protein is predicted to be seryl-tRNA synthetase (serS). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1866(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11789 GB:Z99104 seryl-tRNA synthetase [Bacillus subtilis]
 Identities = 262/425 (61%), Positives = 322/425 (75%), Gaps = 1/425 (0%)

Query: 1   MLDLKRIRTDFDVVAKKLATRGVDQETLTTLKELDIKRRELLIKAEEAKAQRNVASAAIA 60
           MLD K +R +F  +  KL  +G D       + LD +RREL+ K EE K +RN  S  +A
Sbjct: 1   MLDTKMLRANFQEIKAKLVHKGEDLTDFDKFEALDDRRRELIGKVEELKGKRNEVSQQVA 60
```

```
Query: 61  QAKRNKENADEQIAAMQTLSADIKAIDAELADVDANLQSMVTVLPNTPADDVPLGADEDE 120
             KR K++AD  I  M+ +  +IK +D EL  V+A L +++  +PN P + VP+G  ED+
Sbjct: 61  VLKREKKDADHIIKEMREVGEEIKKLDEELRTVEAELDTILLSIPNIPHESVPVGETEDD 120

Query: 121 NVEVRRWGTPREFDFETKAHWDLGESLGILDWERGAKVTGSRFLFYKGLGARLERAIYSF 180
           NVEVR+WG    F +E K HWD+ + LGILD+ER AKVTGSRF+FYKGLGARLERA+Y+F
Sbjct: 121 NVEVRKWGEKPSFAYEPKPHWDIADELGILDFERAAKVTGSRFVFYKGLGARLERALYNF 180

Query: 181 MLDEHAKE-GYTEVIPPYMVNHDSMFGTGQYPKFKEDTFELADSPFVLIPTAEVPLTNYY 239
           MLD H  E  YTEVIPPYMVN  SM GTGQ PKF+ED F++ +  + LIPTAEVP+TN +
Sbjct: 181 MLDLHVDEYNYTEVIPPYMVNRASMTGTGQLPKFEEDAFKIREEDYFLIPTAEVPITNMH 240

Query: 240 RDEIIDGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQELE 299
           RDEI+ G  LPI + A S  FRSEAGSAGRDTRGLIR HQF+KVE+VKF KPE+SY+ELE
Sbjct: 241 RDEILSGDSLPINYAAFSACFRSEAGSAGRDTRGLIRQHQFNKVELVKFVKPEDSYEELE 300

Query: 300 KMTANAENILQKLNLPYRVITLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQ 359
           K+T  AE +LQ L LPYRV+++CTGD+GF+AAK YD+EVWIP+Q+TYREISSCSN E FQ
Sbjct: 301 KLTNQAERVLQLLELPYRVMSMCTGDLGFTAAKKYDIEVWIPSQDTYREISSCSNFEAFQ 360

Query: 360 ARRAQIRYRDEVDGKVRLLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMGNI 419
           ARRA IR+R E  GK   +HTLNGSGLAVGRTVAAILENYQ EDGSV IP+VLRPYMGN
Sbjct: 361 ARRANIRFRREAKGKPEHVHTLNGSGLAVGRTVAAILENYQQEDGSVVIPKVLRPYMGNR 420

Query: 420 DIIKP 424
           +++KP
Sbjct: 421 EVMKP 425
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5225> which encodes the amino acid sequence <SEQ ID 5226>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2453(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 357/424 (84%), Positives = 386/424 (90%)

Query: 1   MLDLKRIRTDFDVVAKKLATRGVDQETLTTLKELDIKRRELLIKAEEAKAQRNVASAAIA 60
           MLDLKRIRTDFD VA KL  RGV ++TLT LKELD KRR LL+++EE KA+RN+ASAAIA
Sbjct: 1   MLDLKRIRTDFDTVAAKLKNRGVSEDTLTHLKELDEKRRALLVQSEELKAERNIASAAIA 60

Query: 61  QAKRNKENADEQIAAMQTLSADIKAIDAELADVDANLQSMVTVLPNTPADDVPLGADEDE 120
           QAKR KE+A +QIA MQ +SADIK ID +L  +D  +  ++TVLPNTP D VP+GADE++
Sbjct: 61  QAKRQKEDATQQIADMQKVSADIKTIDNQLVAIDQQVTDIITVLPNTPHDSVPVGADEED 120

Query: 121 NVEVRRWGTPREFDFETKAHWDLGESLGILDWERGAKVTGSRFLFYKGLGARLERAIYSF 180
           NVE+RRWGTPR+FDFE KAHWDLGE L ILDWERGAKVTG+RFLFYK LGARLERA+Y+F
Sbjct: 121 NVEIRRWGTPRDFDFEVKAHWDLGEDLDILDWERGAKVTGARFLFYKNLGARLERALYNF 180

Query: 181 MLDEHAKEGYTEVIPPYMVNHDSMFGTGQYPKFKEDTFELADSPFVLIPTAEVPLTNYYR 240
           MLDEH KEGY E+I PYMVNHDSMFGTGQYPKFKEDTFELAD+ FVLIPTAEVPLTNYYR
Sbjct: 181 MLDEHIKEGYQEIITPYMVNHDSMFGTGQYPKFKEDTFELADTNFVLIPTAEVPLTNYYR 240

Query: 241 DEIIDGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQELEK 300
            EI+DGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQELEK
Sbjct: 241 GEILDGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQELEK 300

Query: 301 MTANAENILQKLNLPYRVITLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA 360
           MTANAENILQKL LPYRVI+LCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA
Sbjct: 301 MTANAENILQKLGLPYRVISLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA 360
```

```
Query: 361 RRAQIRYRDEVDGKVRLLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMGNID 420
           RRAQIRYRDE DGKV+LLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMG
Sbjct: 361 RRAQIRYRDEADGKVKLLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMGGET 420

Query: 421 IIKP 424
           +I P
Sbjct: 421 VISP 424
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1684

A DNA sequence (GBSx1788) was identified in *S.agalactiae* <SEQ ID 5227> which encodes the amino acid sequence <SEQ ID 5228>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.36   Transmembrane  313 - 329 ( 306 - 332)
   INTEGRAL    Likelihood = -9.24   Transmembrane  159 - 175 ( 155 - 179)
   INTEGRAL    Likelihood = -4.19   Transmembrane   20 -  36 (  16 -  37)
   INTEGRAL    Likelihood = -3.29   Transmembrane  271 - 287 ( 271 - 287)
   INTEGRAL    Likelihood = -2.97   Transmembrane  210 - 226 ( 209 - 227)
   INTEGRAL    Likelihood = -2.87   Transmembrane  242 - 258 ( 241 - 258)
   INTEGRAL    Likelihood = -2.13   Transmembrane   52 -  68 (  50 -  68)

----- Final Results -----
               bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9559> which encodes amino acid sequence <SEQ ID 9560> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA07406 GB:AJ006986 transmembrane protein [Streptococcus pneumoniae]
 Identities = 72/330 (21%), Positives = 143/330 (42%), Gaps = 32/330 (9%)

Query: 14  RHYGLDLLRIISMFMIVITHVLGKGGLRSSVEGHADSYFIVTWIIQVLVYGAVNCYALIS 73
           R+  LDLL++++   +V+ H    GG + +   +  +Y        + ++  VN Y L+
Sbjct: 5   RNINLDLLKVLACVGVVLLHTT-MGGFKETGAWNFLTYLYYLGTYSIPLFFMVNGYLLL- 62

Query: 74  GYVGINSRYRYSKLLSIWAQVFFYTFTITALFAITGHE------VTLLNWRDAFFPIVSG 127
           G   I   Y   K+  +   V  +TF I  LF     E      +  L  +  FF
Sbjct: 63  GKREITYSYILQKIKWLLITVSSWTF-IVWLFKRDFTENLIKKIIGSLIQKGYFF----- 116

Query: 128 QYWYITAYFGLLVFMPVINNGLNALTDKQLKQLVLLMFI--IFSILPAVLNNRVPEFSLS 185
           Q+W+  A   + + +P++   LN+     L  L LLM I  IF +   +L   +  + +
Sbjct: 117 QFWFFGALILIYLCLPILRQFLNS-KRSYLYSLSLLMTIGLIFELSNILLQMPIQTYVIQ 175

Query: 186 KGFEMTWLLILYIIGAYLKRIDL----NIFKTSYLLIIYLLSLVATYAMKFSVGDIW--- 238
                TW    Y++G Y+ +  +    + FK    ++  LL L++   + F    I+
Sbjct: 176 TFRLWTW-FFYYLLGGYIAQFTIEEIESRFKNWMKIVSILLLLISPIILFFIAKTIYHNL 234

Query: 239 ---YWYVSPTLTLGAVSLFILFARASIKPSGFLKKIIVVLAPSTLGVYLCHLHPLIVKYF 295
              Y+Y +  + +  + +F+     ++  +   ++ IV L+  T+GV++  +H  I+K +
Sbjct: 235 FAEYFYDTLFVKVSTLGIFLTILMLTLNEN--RRESIVSLSNQTMGVFI--IHTYIMKVW 290

Query: 296 VRDFAETFVYESIYLYPFLILGAGILIYLL 325
            +     FV   +    F +  + I++ +L
Sbjct: 291 EKVLGFNFVGAYLLFALFTLSVSFIIVGML 320
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1685

A DNA sequence (GBSx1789) was identified in *S.agalactiae* <SEQ ID 5229> which encodes the amino acid sequence <SEQ ID 5230>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2752(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9561> which encodes amino acid sequence <SEQ ID 9562> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46488 GB:AF130465 unknown [Streptococcus salivarius]
 Identities = 88/112 (78%), Positives = 96/112 (85%)

Query: 1    MAQSLNKTVEFQTTGVSYLGMGNKVGKFLVGDQALEFYNDKNVNDYIQIPWTSINQIGAN 60
            MAQSLNKTVE  TTGVSY+ +G KVGKFL+GD ALEFY D NV  YIQIPWTSI QIGAN
Sbjct: 1    MAQSLNKTVELHTTGVSYMAIGGKVGKFLIGDVALEFYPDVNVEQYIQIPWTSITQIGAN 60

Query: 61   VSRKKISRHFEVFTDQGKFLFASKDSGTILKHARRHIGDDKVVKLPTLIQTI 112
            VS K+ISRHFEV TD+ KFLFASKDSG ILK AR H+G++KVVKLPTLIQTI
Sbjct: 61   VSGKRISRHFEVLTDKSKFLFASKDSGKILKIAREHLGNEKVVKLPTLIQTI 112
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5231> which encodes the amino acid sequence <SEQ ID 5232>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3301(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/116 (75%), Positives = 101/116 (87%)

Query: 1    MAQSLNKTVEFQTTGVSYLGMGNKVGKFLVGDQALEFYNDKNVNDYIQIPWTSINQIGAN 60
            MAQSLN +VE++T  VSYLGMG KVG  L+GD+ALEFYNDKNVNDYIQIPWT+IN IGAN
Sbjct: 1    MAQSLNTSVEYKTKAVSYLGMGGKVGHILLGDKALEFYNDKNVNDYIQIPWTAINHIGAN 60

Query: 61   VSRKKISRHFEVFTDQGKFLFASKDSGTILKHARRHIGDDKVVKLPTLIQTILKIF 116
            VSRKK+SRHFE+FTDQGKFLFAS DSG ILK  R+HIG++KV+ LPTL+QT +  F
Sbjct: 61   VSRKKVSRHFEIFTDQGKFLFASGDSGKILKITRQHIGNEKVITLPTLMQTFINKF 116
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1686

A DNA sequence (GBSx1790) was identified in *S.agalactiae* <SEQ ID 5233> which encodes the amino acid sequence <SEQ ID 5234>. This protein is predicted to be mannose-specific phosphotransferase system component IID (manZ). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.92   Transmembrane  281 - 297 ( 279 - 302)
   INTEGRAL    Likelihood = -4.88   Transmembrane  187 - 203 ( 185 - 205)
   INTEGRAL    Likelihood = -4.35   Transmembrane  260 - 276 ( 257 - 277)
   INTEGRAL    Likelihood = -1.01   Transmembrane  129 - 145 ( 129 - 145)

----- Final Results -----
               bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46487 GB:AF130465 mannose-specific phosphotransferase system
           component IID [Streptococcus salivarius]
 Identities = 247/303 (81%), Positives = 276/303 (90%)

Query: 1   MTEQIKLSKSDRQKVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTTKEDRAAALER 60
           M E+I+LS++DR+KVWWRSQFLQGSWNYERMQN+GWAY+LIPA+KKLYT KED+AAAL+R
Sbjct: 1   MAEKIQLSQADRKKVWWRSQFLQGSWNYERMQNLGWAYSLIPAIKKLYTNKEDQAAALKR 60

Query: 61  HMEFFNTHPYVAAPIIGVTLALEEEKASGTPVEDKAIQGVKIGMMGPLAGIGDPVFWFTV 120
           H+EFFNTHPYVAAPI+GVTLALEEEKA+GT +ED AIQGVKIGMMGPLAGIGDPVFWFTV
Sbjct: 61  HLEFFNTHPYVAAPIMGVTLALEEEKANGTDIEDAAIQGVKIGMMGPLAGIGDPVFWFTV 120

Query: 121 RPILGALGASLASAGNILGPIIFFVGWNLIRMSFLWYTQELGYKSGKEITKDMSGGILQD 180
           RPILGALGASLA AGNI GP+IFF+GWNLIRM+FLWYTQELGYK+G EITKDMSGGIL+D
Sbjct: 121 RPILGALGASLAQAGNIAGPLIFFIGWNLIRMAFLWYTQELGYKAGSEITKDMSGGILKD 180

Query: 181 ITKGASILGMFILAVLVKRWVAINFTVDLPKKTLSEGAYINFPKDHVSGQQLHDILGQVQ 240
           ITKGASILGMFILAVLV+RWV+I FTV+LP K LS+GAYI +PK +VSG QL  ILGQV
Sbjct: 181 ITKGASILGMFILAVLVERWVSIVFTVNLPGKVLSKGAYIEWPKGNVSGDQLKTILGQVN 240

Query: 241 SGLSLDKMQPQTLQGQLDSLIPGLAGLLLTFFCMWLLKKKVSPITIIIGLFIVGILARLA 300
             LS DK+Q  TLQ QLDSLIPGL GLLLTF CMWLLKKKVSPITIIIGLF+VGI+A
Sbjct: 241 DKLSFDKIQVDTLQKQLDSLIPGLMGLLLTFACMWLLKKKVSPITIIIGLFVVGIVASFF 300

Query: 301 GVM 303
           G+M
Sbjct: 301 GIM 303
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5235> which encodes the amino acid sequence <SEQ ID 5236>. Analysis of this protein sequence reveals the following:

```
     Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.39   Transmembrane  284 - 300 ( 279 - 302)
   INTEGRAL    Likelihood = -4.88   Transmembrane  261 - 277 ( 257 - 278)
   INTEGRAL    Likelihood = -4.51   Transmembrane  181 - 197 ( 180 - 198)

----- Final Results -----
               bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD46487 GB:AF130465 mannose-specific phosphotransferase system
           component IID [Streptococcus salivarius]
```

```
 Identities = 239/303 (78%), Positives = 268/303 (87%)

Query: 1    MTEQIKLTKSDRQRVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTSPEDRAAALER 60
            M E+I+L+++DR++VWWRSQFLQGSWNYERMQN+GWAY+LIPA+KKLYT+ ED+AAAL+R
Sbjct: 1    MAEKIQLSQADRKKVWWRSQFLQGSWNYERMQNLGWAYSLIPAIKKLYTNKEDQAAALKR 60

Query: 61   HMEFFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKIGMMGPLAGIGDPVFWFTI 120
            H+EFFNTHPYVAAPI+GVTLALEEE+ANGT I+D AIQGVKIGMMGPLAGIGDPVFWFT+
Sbjct: 61   HLEFFNTHPYVAAPIMGVTLALEEEKANGTDIEDAAIQGVKIGMMGPLAGIGDPVFWFTV 120

Query: 121  RPILGALGASLASTGNIVGPLLFFFGWNLIRMAFLWYTQEFGYKAGSEITKDMSGGILQD 180
            RPILGALGASLA  GNI GPL+FF GWNLIRMAFLWYTQE GYKAGSEITKDMSGGIL+D
Sbjct: 121  RPILGALGASLAQAGNIAGPLIFFIGWNLIRMAFLWYTQELGYKAGSEITKDMSGGILKD 180

Query: 181  ITKGASILGMFILAVLVQRWVSINFTIDLPGKQLSDGAYVVFPDGAVKGAELKTILANAI 240
            ITKGASILGMFILAVLV+RWVSI FT++LPGK LS GAY+ +P G V G +LKTIL
Sbjct: 181  ITKGASILGMFILAVLVERWVSIVFTVNLPGKVLSKGAYIEWPKGNVSGDQLKTILGQVN 240

Query: 241  GGMSLDKVQAQTLQGQLDSLIPGLAGLLLTFLCMWLLKKKVSPIAIIIGLFAFGILAHLA 300
              +S DK+Q  TLQ QLDSLIPGL GLLLTF CMWLLKKKVSPI IIIGLF  GI+A
Sbjct: 241  DKLSFDKIQVDTLQKQLDSLIPGLMGLLLTFACMWLLKKKVSPITIIIGLFVVGIVASFF 300

Query: 301  GIM 303
            GIM
Sbjct: 301  GIM 303
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 255/303 (84%), Positives = 277/303 (91%)

Query: 1    MTEQIKLSKSDRQKVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTTKEDRAAALER 60
            MTEQIKL+KSDRQ+VWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYT+ EDRAAALER
Sbjct: 1    MTEQIKLTKSDRQRVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTSPEDRAAALER 60

Query: 61   HMEFFNTHPYVAAPIIGVTLALEEEKASGTPVEDKAIQGVKIGMMGPLAGIGDPVFWFTV 120
            HMEFFNTHPYVAAPIIGVTLALEEE+A+GTP++DKAIQGVKIGMMGPLAGIGDPVFWFT+
Sbjct: 61   HMEFFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKIGMMGPLAGIGDPVFWFTI 120

Query: 121  RPILGALGASLASAGNILGPIIFFVGWNLIRMSFLWYTQELGYKSGKEITKDMSGGILQD 180
            RPILGALGASLAS GNI+GP++FF GWNLIRM+FLWYTQE GYK+G EITKDMSGGILQD
Sbjct: 121  RPILGALGASLASTGNIVGPLLFFFGWNLIRMAFLWYTQEFGYKAGSEITKDMSGGILQD 180

Query: 181  ITKGASILGMFILAVLVKRWVAINFTVDLPKKTLSEGAYINFPKDHVSGQQLHDILGQVQ 240
            ITKGASILGMFILAVLV+RWV+INFT+DLP K LS+GAY+ FP   V G +L  IL
Sbjct: 181  ITKGASILGMFILAVLVQRWVSINFTIDLPGKQLSDGAYVVFPDGAVKGAELKTILANAI 240

Query: 241  SGLSLDKMQPQTLQGQLDSLIPGLAGLLLTFFCMWLLKKKVSPITIIIGLFIVGILARLA 300
             G+SLDK+Q QTLQGQLDSLIPGLAGLLLTF CMWLLKKKVSPI IIIGLF  GILA LA
Sbjct: 241  GGMSLDKVQAQTLQGQLDSLIPGLAGLLLTFLCMWLLKKKVSPIAIIIGLFAFGILAHLA 300

Query: 301  GVM 303
            G+M
Sbjct: 301  GIM 303
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1687

A DNA sequence (GBSx1791) was identified in *S.agalactiae* <SEQ ID 5237> which encodes the amino acid sequence <SEQ ID 5238>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
              bacterial cytoplasm --- Certainty=0.2580(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1688

A DNA sequence (GBSx1792) was identified in *S.agalactiae* <SEQ ID 5239> which encodes the amino acid sequence <SEQ ID 5240>. This protein is predicted to be mannose-specific phosphotransferase system component IIC (manY). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.95    Transmembrane  142 - 158 ( 137 - 165)
   INTEGRAL    Likelihood = -2.60    Transmembrane   65 -  81 (  61 -  81)
   INTEGRAL    Likelihood = -1.97    Transmembrane  103 - 119 ( 103 - 122)

----- Final Results -----
               bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9301> which encodes amino acid sequence <SEQ ID 9302> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46486 GB:AF130465 mannose-specific phosphotransferase system
           component IIC [Streptococcus salivarius]
 Identities = 134/186 (72%), Positives = 154/186 (82%), Gaps = 1/186 (0%)

Query: 1   MVKSGDFTQKGINFAFSTAVPLAIAGLFLTMIVRTISTALVHAGDKAASEGNFAAIERFH 60
           +VK G+FT +GI  A +TA+PLA+AGLFLTM+VRT S ALVHA DKAA  GN A +ER H
Sbjct: 86  LVKGGNFTTEGIGVATATAIPLAVAGLFLTMLVRTASVALVHAADKAAESGNIAGVERAH 145

Query: 61  FIALLLQGLRIAFPAALLLAIPSSSVQSILEAMPDWLNGGMQVGGAMVVAVGYAMVINMM 120
           ++ALLLQGLRIA PAALLLAIP+ SVQ  L  MP WLN GM VGG MVVAVGYAMVINMM
Sbjct: 146 YLALLLQGLRIAVPAALLLAIPAESVQHALGLMPSWLNHGMVVGGGMVVAVGYAMVINMM 205

Query: 121 ATREVWPFFALGFALAALNQLTLIAMGTIGVAIALIYISLSKMGGSK-GTSNAGSNDPIG 179
           ATREVWPFFA+GFA AA++QLTLIA+G IGVAIA IY++LSK GG   G +++GS DPIG
Sbjct: 206 ATREVWPFFAIGFAFAAISQLTLIALGAIGVAIAFIYLNLSKQGGGNGGGTSSGSGDPIG 265

Query: 180 DILEDY 185
           DILEDY
Sbjct: 266 DILEDY 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5241> which encodes the amino acid sequence <SEQ ID 5242>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30    Transmembrane    4 -  20 (   1 -  28)
   INTEGRAL    Likelihood = -7.64    Transmembrane  226 - 242 ( 212 - 247)
   INTEGRAL    Likelihood = -4.14    Transmembrane  102 - 118 ( 101 - 123)
   INTEGRAL    Likelihood = -3.77    Transmembrane   71 -  87 (  69 -  87)
   INTEGRAL    Likelihood = -3.40    Transmembrane  150 - 166 ( 146 - 167)
```

```
   INTEGRAL    Likelihood = -2.13   Transmembrane  186 - 202 ( 186 - 202)
   INTEGRAL    Likelihood = -0.37   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
              bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD46486 GB:AF130465 mannose-specific phosphotransferase system
           component IIC [Streptococcus salivarius]
 Identities = 211/271 (77%), Positives = 237/271 (86%), Gaps = 2/271 (0%)

Query: 1   MSDISIISAILVVIIAFFAGLEGILDQFQMHQPLVACTLIGLVTGHLEAGVILGGTLQML 60
           MSD+SIISAILVV++AF AGLEGILDQFQ HQPLVACTLIG  TG+L AG++LGG+LQM+
Sbjct: 1   MSDMSIISAILVVVVAFLAGLEGILDQFQFHQPLVACTLIGAATGNLTAGIMLGGSLQMI 60

Query: 61  ALGWANIGAAVAPDAALASVAAAIIMVKSGDFTQKGITFAYSTAIPLAVAGLFLTMIVRT 120
           AL WANIGAAVAPDAALASVAAAII+VK G+FT +GI  A +TAIPLAVAGLFLTM+VRT
Sbjct: 61  ALAWANIGAAVAPDAALASVAAAIILVKGGNFTTEGIGVATATAIPLAVAGLFLTMLVRT 120

Query: 121 LSTALVHAGDKAAAEGNFAGIERFHFIALLLQGLRIAVPAALLVAVPTSAVQSVLNAMPN 180
            S ALVHA DKAA  GN AG+ER H++ALLLQGLRIAVPAALL+A+P  +VQ  L  MP+
Sbjct: 121 ASVALVHAADKAAESGNIAGVERAHYLALLLQGLRIAVPAALLLAIPAESVQHALGLMPS 180

Query: 181 WLNEGMQIGGAMVVAVGYAMVINMMATREVWPFFALGFALAAISQLTLIAMGVIGVAIAF 240
           WLN GM +GG MVVAVGYAMVINMMATREVWPFFA+GFA AAISQLTLIA+G IGVAIAF
Sbjct: 181 WLNHGMVVGGGMVVAVGYAMVINMMATREVWPFFAIGFAFAAISQLTLIALGAIGVAIAF 240

Query: 241 IYLNLSKKGG--NGGNAAGSADPIGDILEDY 269
           IYLNLSK+GG   GG ++GS DPIGDILEDY
Sbjct: 241 IYLNLSKQGGGNGGTSSGSGDPIGDILEDY 271
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 155/185 (83%), Positives = 173/185 (92%), Gaps = 1/185 (0%)

Query: 1   MVKSGDFTQKGINFAFSTAVPLAIAGLFLTMIVRTISTALVHAGDKAASEGNFAAIERFH 60
           MVKSGDFTQKGI FA+STA+PLA+AGLFLTMIVRT+STALVHAGDKAA+EGNFA IERFH
Sbjct: 86  MVKSGDFTQKGITFAYSTAIPLAVAGLFLTMIVRTLSTALVHAGDKAAAEGNFAGIERFH 145

Query: 61  FIALLLQGLRIAFPAALLLAIPSSSVQSILEAMPDWLNGGMQVGGAMVVAVGYAMVINMM 120
           FIALLLQGLRIA PAALL+A+P+S+VQS+L AMP+WLN GMQ+GGAMVVAVGYAMVINMM
Sbjct: 146 FIALLLQGLRIAVPAALLVAVPTSAVQSVLNAMPNWLNEGMQIGGAMVVAVGYAMVINMM 205

Query: 121 ATREVWPFFALGFALAALNQLTLIAMGTIGVAIALIYISLSKMGGSKGTSNAGSNDPIGD 180
           ATREVWPFFALGFALAA++QLTLIAMG IGVAIA IY++LSK GG+ G + AGS DPIGD
Sbjct: 206 ATREVWPFFALGFALAAISQLTLIAMGVIGVAIAFIYLNLSKKGGNGGNA-AGSADPIGD 264

Query: 181 ILEDY 185
           ILEDY
Sbjct: 265 ILEDY 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1689

A DNA sequence (GBSx1793) was identified in *S.agalactiae* <SEQ ID 5243> which encodes the amino acid sequence <SEQ ID 5244>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
               bacterial cytoplasm --- Certainty=0.3171(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1690

A DNA sequence (GBSx1794) was identified in *S.agalactiae* <SEQ ID 5245> which encodes the amino acid sequence <SEQ ID 5246>. This protein is predicted to be pseudouridine synthase (rluC). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2717(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06566 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 124/281 (44%), Positives = 171/281 (60%), Gaps = 8/281 (2%)

Query: 16  LLKSHDVSRGLLAKIKYRGGKIFVNGEEQNAIFLLEIGDVVTIDIPDE-PSHETL-EPVP 73
            L +   VS+  LA IK++GG I +NGEE      + + D VT+++P E PS   + EPVP
Sbjct: 24  LREGKHVSKRSLAAIKFKGGTILLNGEEVTVRETVHVNDQVTLELPHEYPSPSMIAEPVP 83

Query: 74  HDLDIIYEDDHFLILNKPFGFASIPSSIH-SNTIANFIKHYYVSNNYANQQVHIVTRLDR 132
               D+IYE+DH+L++NKP G  +IPS  H   T+AN + +Y+     A    H V RLD+
Sbjct: 84  --FDVIYENDHYLVVNKPAGVPTIPSRDHPQGTLANGLLNYFQRQKMA-ATFHAVNRLDK 140

Query: 133 DTSGLMLFAKHGYAHARLDKQLQAKAIEKRYYALVSGSGDLADSGDIIAPIARDVDSIIT 192
           DTSGL++ AKH  AH +L KQ +   I++ Y A+V G  +  + G I APIAR  +S+IT
Sbjct: 141 DTSGLLIVAKHQLAHDQLSKQQRQGNIKRTYMAIVQGEIEQQE-GTITAPIARKEESLIT 199

Query: 193 RRVHESGKYAHTSYQVVARYGDVRLVDIKLHTGRTHQIRVHFAHIGFPLLGDDLYGGRMD 252
           R V E G+ A T ++V+ R     +V ++L TGRTHQIRVHF+++G+PL GDDLYGG
Sbjct: 200 REVREDGQLAITHFKVIDRLNQGTIVQVQLETGRTHQIRVHFSYLGYPLFGDDLYGGERK 259

Query: 253 LGINRQALHCHSLSFYDPFMGKINKQTLDLTDDFDSVIMEL 293
            GI RQALH   L+ + PF       T  L  D   +I  L
Sbjct: 260 -GIERQALHSTELTIHCPFTEVEQTFTEGLPPDMKELIRHL 299
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5247> which encodes the amino acid sequence <SEQ ID 5248>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2786(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 223/294 (75%), Positives = 251/294 (84%), Gaps = 1/294 (0%)
```

```
Query: 1   MKFEYVAKERCKVKTLLKSHDVSRGLLAKIKYRGGKIFVNGEEQNAIFLLEIGDVVTIDI 60
           M+FE+VA +R KVKTLLKS+DVS+GLLAKIKY+GG I VNG EQNAI+LL++GDVVTIDI
Sbjct: 1   MRFEFVADKRIKVKTLLKSYDVSKGLLAKIKYKGGNILVNGIEQNAIYLLQVGDVVTIDI 60

Query: 61  PDEPSHETLEPVPHDLDIIYEDDHFLILNKPFGFASIPSSIHSNTIANFIKHYYVSNNYA 120
           P+E   E LE +P DLDI++EDDHFL++NKP GFASIPS+IHSNTIANFIK YYV N+Y
Sbjct: 61  PNEEPFEKLEAIPFDLDIVHEDDHFLVINKPIGFASIPSAIHSNTIANFIKAYYVDNHYL 120

Query: 121 NQQVHIVTRLDRDTSGLMLFAKHGYAHARLDKQLQAKAIEKRYYALVSGSGDLADSGDII 180
           +QQVHIVTRLDRDTSGLMLFAKHGYAHARLDKQLQ ++IEKRY+ALVSG+G L D GDII
Sbjct: 121 DQQVHIVTRLDRDTSGLMLFAKHGYAHARLDKQLQTRSIEKRYFALVSGNGMLPDEGDII 180

Query: 181 APIARDVDSIITRRVHESGKYAHTSYQVVARYGD-VRLVDIKLHTGRTHQIRVHFAHIGF 239
           API R  DSIITR V   GKYA TSY+VVARY + V LVDIKLHTGRTHQIRVHFAHIGF
Sbjct: 181 APIGRSKDSIITRAVDPMGKYAKTSYKVVARYSENVHLVDIKLHTGRTHQIRVHFAHIGF 240

Query: 240 PLLGDDLYGGRMDLGINRQALHCHSLSFYDPFMGKINKQTLDLTDDFDSVIMEL 293
           PLLGDDLYGGR+DLGI RQALHCH L+F DPF        + LTDDFDSVI+ L
Sbjct: 241 PLLGDDLYGGRLDLGITRQALHCHYLNFKDPFTESDCSYAIHLTDDFDSVIIGL 294
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1691

A DNA sequence (GBSx1795) was identified in *S.agalactiae* <SEQ ID 5249> which encodes the amino acid sequence <SEQ ID 5250>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1521(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9845> which encodes amino acid sequence <SEQ ID 9846> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13018 GB:Z99110 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 120/267 (44%), Positives = 174/267 (64%), Gaps = 3/267 (1%)

Query: 13  RVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGMLLSAFHMYEKQLD 72
           + A+ + G   S  + SK+  A+  D D  L + +P+IVIS+GGDG LL AFH Y  +LD
Sbjct: 2   KFAVSSKGDQVSDTLKSKI-QAYLLDFDMELDENEPEIVISVGGDGTLLYAFHRYSDRLD 60

Query: 73  KVRFVGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTITL-EDGRVIRARA 131
           K  FVGVHTGHLGFY D+   E++ L+  +       + YP+L+V +T  E+ R  R  A
Sbjct: 61  KTAFVGVHTGHLGFYADWVPHEIEKLVLAIAKTPYHTVEYPLLEVIVTYHENEREERYLA 120

Query: 132 LNESTIKRIEKTMVADVVINQVVFERFRGDGILVSTPTGSTAYNKSLGGAVLHPTIEALQ 191
           LNE TIK IE ++VADV I   +FE FRGDG+ +STP+GSTAYNK+LGGA++HP+I A+Q
Sbjct: 121 LNECTIKSIEGSLVADVEIKGQLFETFRGDGLCLSTPSGSTAYNKALGGAIIHPSIRAIQ 180

Query: 192 LTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVYTISIDNKTVHYKNVTKIEYSIDE 251
           L E++S+NNRV+RT+GS +++P      I P+    + ++ID+ T+ +K+V  I   +
Sbjct: 181 LAEMASINNRVFRTVGSPLLLPSHHDCMIKPRNEVDFQVTIDHLTLLHKDVKSIRCQVAS 240

Query: 252 KSINFVSTPSHTSFWERVNDAFIGEPE 278
           + + F        FW+RV D+FIG+ E
Sbjct: 241 EKVRFARFRPF-PFWKRVQDSFIGKGE 266
```

A related sequence was also identified in GAS <SEQ ID 9137> which encodes the amino acid sequence <SEQ ID 9138>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2190(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 155-157
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 232/276 (84%), Positives = 257/276 (93%)

Query: 1   MMTQMNFTDRATRVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGML 60
            +MTQMN+T +  RVAIIANGKYQSKRVASKLF+ FK DPDFYLSKK+PDIVISIGGDGML
Sbjct: 1   VMTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGML 60

Query: 61  LSAFHMYEKQLDKVRFVGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTIT 120
            LSAFHMYEK+LDKVRFVG+HTGHLGFYTDYRDFEVD LI+NL+ DKGEQISYPILKV IT
Sbjct: 61  LSAFHMYEKELDKVRFVGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAIT 120

Query: 121 LEDGRVIRARALNESTIKRIEKTMVADVVINQVVFERFRGDGILVSTPTGSTAYNKSLGG 180
            L+DGRV++ARALNE+T+KRIEKTMVADV+IN V FE FRGDGI VSTPTGSTAYNKSLGG
Sbjct: 121 LDDGRVVKARALNEATVKRIEKTMVADVIINHVKFESFRGDGISVSTPTGSTAYNKSLGG 180

Query: 181 AVLHPTIEALQLTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVYTISIDNKTVHYK 240
            AVLHPTIEALQLTEISSLNNRV+RTLGSS+IIPKKD IE+VPKR+G+YTISIDNKT   K
Sbjct: 181 AVLHPTIEALQLTEISSLNNRVFRTLGSSIIIPKKDKIELVPKRLGIYTISIDNKTYQLK 240

Query: 241 NVTKIEYSIDEKSINFVSTPSHTSFWERVNDAFIGE 276
            NVTK+EY ID++ I+FVS+PSHTSFWERV DAFIGE
Sbjct: 241 NVTKVEYFIDDEKIHFVSSPSHTSFWERVKDAFIGE 276
```

A related GBS gene <SEQ ID 8879> and protein <SEQ ID 8880> were also identified. Analysis of this protein sequence reveals an RGD motif at residues 159-161.

The protein has homology with the following sequences in the databases:

```
45.0/65.6% over 264aa
                                                           Bacillus subtilis
   EGAD|107338| hypothetical protein Insert characterized OMNI|NT01BS1363 BC541A protein-
related Insert characterized
  SP|O31612|YJBN_BACSU HYPOTHETICAL 30.0 KDA PROTEIN IN MECA-TENA INTERGENIC REGION. Insert
characterized
  GP|2633515|emb|CAB13018.1||Z99110 similar to hypothetical proteins Insert characterized
  PIR|F69844|F69844 conserved hypothetical protein yjbN - Insert characterized

ORF02026(337 - 1134 of 1437)
EGAD|107338|BS1162(2 - 266 of 266) hypothetical protein {Bacillus subtilis} OMNI|NT01BS1363
BC541A protein-related SP|O31612|YJBN_BACSU HYPOTHETICAL 30.0 KDA PROTEIN IN MECA-TENA
INTERGENIC REGION. GP|2633515|emb|CAB13018.1||Z99110 similar to hypothetical proteins
{Bacillus subtilis} PIR|F69844|F69844 conserved hypothetical protein yjbN - Bacillus
subtilis
%Match = 22.8
%Identity = 44.9  %Similarity = 65.5
Matches = 120  Mismatches = 89  Conservative Sub.s = 55

87        117       147       177       207       237       267       297
RKF*QKYKSELWL*IFGQPSNIH*ITSIRGTSLKKLNKDWRKQQKSL*NWMKKCVRFAKIFVKHSFYLIL*IEN*AMV*E

327       357       387       417       447       477       507       537
```

```
IVMTQMNFTDRATRVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGMLLSAFHMYEKQLDKVRFVGV
             : |: : |   |  : ||: | :  | |  | : :|:||||:|||| || ||| |  :|||  ||||
            MKFAVSSKGDQVSDTLKSKIQA-YLLDFDMELDENEPEIVISVGGDGTLLYAFHRYSDRLDKTAFVGV
                    10        20        30        40        50        60

567       597       627       657       684       714       744       774
HTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTITL-EDGRVIRARALNESTIKRIEKTMVADVVINQVVFERF
|||||||| |:   |:: |:  :       : ||:|:| :|  |: |  |  |||| ||| || ::|||| |   :|| |
HTGHLGFYADWVPHEIEKLVLAIAKTPYHTVEYPLLEVIVTYHENEREERYLALNECTIKSIEGSLVADVEIKGQLFETF
          80        90       100       110       120       130       140

804       834       864       894       924       954       984       1014
RGDGILVSTPTGSTAYNKSLGGAVLHPTIEALQLTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVYTISIDNKTVH
||||: :|||:|||||||:||||::||:| |:|| |::|:||||:||:|| :::|      | |:    : ::||: |:
RGDGLCLSTPSGSTAYNKALGGAIIHPSIRAIQLAEMASINNRVFRTVGSPLLLPSHHDCMIKPRNEVDFQVTIDHLTLL
         160       170       180       190       200       210       220

1044      1074      1104      1134      1164      1194      1224      1254
YKNVTKIEYSIDEKSINFVSTPSHTSFWERVNDAFIGEPEH*NLNT*QKKGAKLKHF*KVMMFQGGY*QRLSTEVVRFLL
:|:|  |   :  : : | :      ||:|| |:|||: |
HKDVKSIRCQVASEKVRF-ARFRPFPFWKRVQDSFIGKGE
         240       250       260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5251> which encodes the amino acid sequence <SEQ ID 5252>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2190(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  481 bits (1224), Expect = e-138
 Identities = 233/276 (84%), Positives = 257/276 (92%)

Query: 1   VMTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGML 60
            VMTQMN+T +  RVAIIANGKYQSKRVASKLF+ FK DPDFYLSKK+PDIVISIGGDGML
Sbjct: 1   VMTQMNFTDRATRVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGML 60

Query: 61  LSAFHMYEKELDKVRFVGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAIT 120
            LSAFHMYEK+LDKVRFVG+HTGHLGFYTDYRDFEVD LI+NL+ DKGEQISYPILKV IT
Sbjct: 61  LSAFHMYEKQLDKVRFVGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTIT 120

Query: 121 LDDGRVVKARALNEATVKRIEKTMVADVIINHVKFESFRGDGISVSTPTGSTAYNKSLGG 180
            L+DGRV++ARALNE+T+KRIEKTMVADV+IN V FE FRGDGI VSTPTGSTAYNKSLGG
Sbjct: 121 LEDGRVIRARALNESTIKRIEKTMVADVVINQVVFERFRGDGILVSTPTGSTAYNKSLGG 180

Query: 181 AVLHPTIEALQLTEISSLNNRVFRTLGSSIIIPKKDKIELVPKRLGIYTISIDNKTYQLK 240
            AVLHPTIEALQLTEISSLNNRV+RTLGSS+IIPKKD IE+VPKR+G+YTISIDNKT   K
Sbjct: 181 AVLHPTIEALQLTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVYTISIDNKTVHYK 240

Query: 241 NVTKVEYFIDDEKIHFVSSPSHTSFWERVKDAFIGE 276
            NVTK+EY ID++ I+FVS+PSHTSFWERV DAFIGE
Sbjct: 241 NVTKIEYSIDEKSINFVSTPSHTSFWERVNDAFIGE 276
```

SEQ ID 8880 (GBS308) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 3; MW 59kDa).

GBS308-GST was purified as shown in Figure 226, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1692

A DNA sequence (GBSx1796) was identified in *S.agalactiae* <SEQ ID 5253> which encodes the amino acid sequence <SEQ ID 5254>. This protein is predicted to be permease. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3653(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06568 GB:AP001516 GTP pyrophosphokinase [Bacillus halodurans]
 Identities = 115/208 (55%), Positives = 159/208 (76%), Gaps = 3/208 (1%)

Query: 4   DWETFLDPYIQTVGELKIKLRGIRKQFRKQNRHSPIEFVTGRVKSVESIQEKMVLRGISE 63
           +W+ FL PY Q V ELK+KL+GIR+Q++K ++H+PIEFVTGRVK + SI +K + + I
Sbjct: 3   NWDVFLTPYKQAVEELKVKLKGIREQYQKSSKHTPIEFVTGRVKPISSILDKAIRKNIPL 62

Query: 64  ENLAQDLQDIAGLRIMVQFVDDVDEVLALLRKRHDMTVVQERDYITHMKSSGYRSYHVVV 123
           + L + +QD+AGLRI+ QFV+D++ V+ L+R R D  +V+ERDY+   K SGYRSYH+V+
Sbjct: 63  DQLEEKMQDLAGLRIVTQFVEDIETVVQLIRSRSDFEIVEERDYVEQKKDSGYRSYHLVL 122

Query: 124 EYPVDTIDGQKKVLAEIQIRTLAMNFWATIEHSLNYKYQGDFPEEIKQRLEKTAKIALEL 183
            YPV TI+G+K++L E+QIRTLAMNFWATIEHSLNYKY G+ P  IK RL++ A+ A  L
Sbjct: 123 RYPVQTIEGEKRILVELQIRTLAMNFWATIEHSLNYKYSGEIPLNIKTRLQRAAEAAFRL 182

Query: 184 DEEMRKIREDIREAQLLFDPLNRKLSDG 211
           DEEM +IR+++REAQ +   + RK   G
Sbjct: 183 DEEMSQIRDEVREAQQI---ITRKQEQG 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5255> which encodes the amino acid sequence <SEQ ID 5256>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4064(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 196/223 (87%), Positives = 213/223 (94%)

Query: 1   MSMDWETFLDPYIQTVGELKIKLRGIRKQFRKQNRHSPIEFVTGRVKSVESIQEKMVLRG 60
           M++DWE FLDPYIQTVGELKIKLRGIRKQ+RKQNR+SPIEFVTGRVKS+ESI+EKM+LRG
Sbjct: 1   MTLDWEEFLDPYIQTVGELKIKLRGIRKQYRKQNRYSPIEFVTGRVKSIESIKEKMILRG 60

Query: 61  ISEENLAQDLQDIAGLRIMVQFVDDVDEVLALLRKRHDMTVVQERDYITHMKSSGYRSYH 120
           + EEN+AQD+QDIAGLRIMVQFVDDV+EVLALLR+R DMT+V ERDYI +MKSSGYRSYH
Sbjct: 61  VIEENIAQDIQDIAGLRIMVQFVDDVEEVLALLRQRQDMTIVYERDYIRNMKSSGYRSYH 120

Query: 121 VVVEYPVDTIDGQKKVLAEIQIRTLAMNFWATIEHSLNYKYQGDFPEEIKQRLEKTAKIA 180
           VVVEYPVDTI+GQKKVLAEIQIRTLAMNFWATIEHSLNYKY GDFPEEIK+RLE TAKIA
Sbjct: 121 VVVEYPVDTIEGQKKVLAEIQIRTLAMNFWATIEHSLNYKYGGDFPEEIKKRLEVTAKIA 180
```

```
Query: 181 LELDEEMRKIREDIREAQLLFDPLNRKLSDGVGNSDDTDEFYR 223
           LELDEEMRKIREDIREAQLLFDP+ R LSDGVGNSDDTDE YR
Sbjct: 181 LELDEEMRKIREDIREAQLLFDPVTRNLSDGVGNSDDTDELYR 223
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1693

A DNA sequence (GBSx1797) was identified in *S.agalactiae* <SEQ ID 5257> which encodes the amino acid sequence <SEQ ID 5258>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13015 GB:Z99110 yjbK [Bacillus subtilis]
 Identities = 63/184 (34%), Positives = 99/184 (53%), Gaps = 10/184 (5%)

Query: 4   LEIEYKTLLNKDEFNRLTSLFSHVQP--ITQTNYYFDTETFEMKAHRMSLRIRTLPNRAE 61
           +EIE+K +L K EF  + S     +     Q N+YFDT++F +K    +LRIR    +
Sbjct: 5   IEIEFKNMLTKQEFKNIASALQLTEKDFTDQKNHYFDTDSFALKQKHAALRIRRKNGKYV 64

Query: 62  LTLKIPREVGNLEHNHDLT--LEEAKYIVKNGQFPEDTEIASLILEKGVDPTKLAVFGQL 119
           LTLK P +VG LE +  L+   + A + V  G  P   ++  L     +D   +  FG L
Sbjct: 65  LTLKEPADVGLLETHQQLSEVSDLAGFSVPEG--PVKDQLHKL----QIDTDAIQYFGSL 118

Query: 120 TTTRREMETSIGLMALDSNIYADIKDYELELEVKQPKQGKRDFDQFLKENNINFKYAKSK 179
            T R E ET  GL+ LD + Y + +DYE+E E     +G++ F++ L++ +I  +  K+K
Sbjct: 119 ATNRAEKETEKGLIVLDHSRYLNKEDYEIEFEAADWHEGRQAFEKLLQQFSIPQRETKNK 178

Query: 180 VARF 183
           + RF
Sbjct: 179 ILRF 182
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5259> which encodes the amino acid sequence <SEQ ID 5260>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3470(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/188 (60%), Positives = 139/188 (73%), Gaps = 1/188 (0%)

Query: 1   MTHLEIEYKTLLNKDEFNRLTSLFSHVQPITQTNYYFDTETFEMKAHRMSLRIRTLPNRA 60
           MT+LEIEYKTLL K+E+NRL S   HV P+TQTNYY DT+ F++KA++MSLRIRT  N A
Sbjct: 1   MTNLEIEYKTLLTKNEYNRLLSQMKHVTPVTQTNYYIDTKAFDLKANKMSLRIRTFVNSA 60

Query: 61  ELTLKIPREVGNLEHNHDLTLEEAKYIVKNGQFPEDTEIASLILEKGVDPTKLAVFGQLT 120
           ELTLK+P +VGN E+N  L LE+AK ++K+G  PE T +  +I+ KG+ P+ L  FG LT
Sbjct: 61  ELTLKVPEKVGNREYNVPLFLEQAKDMIKHGNLPESTAL-DIIISKGIKPSALVTFGNLT 119
```

```
Query: 121 TTRREMETSIGLMALDSNIYADIKDYELELEVKQPKQGKRDFDQFLKENNINFKYAKSKV 180
           T RRE    IG +ALD N+YA+ KDYELELEV    QGK DFD FL E +I FKYAKSKV
Sbjct: 120 TVRRETVIPIGKLALDYNLYANTKDYELELEVSDALQGKIDFDSFLSEYHITFKYAKSKV 179

Query: 181 ARFSATLK 188
           AR   TLK
Sbjct: 180 ARCINTLK 187
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1694

A DNA sequence (GBSx1798) was identified in *S.agalactiae* <SEQ ID 5261> which encodes the amino acid sequence <SEQ ID 5262>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1815(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1695

A DNA sequence (GBSx1799) was identified in *S.agalactiae* <SEQ ID 5263> which encodes the amino acid sequence <SEQ ID 5264>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0621(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1696

A DNA sequence (GBSx1800) was identified in *S.agalactiae* <SEQ ID 5265> which encodes the amino acid sequence <SEQ ID 5266>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
              bacterial cytoplasm --- Certainty=0.3369(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11827 GB:Z99104 phosphoribosyl pyrophosphate synthetase
          [Bacillus subtilis]
 Identities = 166/319 (52%), Positives = 231/319 (72%), Gaps = 4/319 (1%)

Query: 1   MAEQYADKQIKLFSLTANREIAEKISQASGIPLGKMSSRQFSDGEIMINIEETVRGDDIY 60
           M+ QY DK +K+FSL +N E+A++I+   G+ LGK S  +FSDGE+ INIEE++RG D Y
Sbjct: 1   MSNQYGDKNLKIFSLNSNPELAKEIADIVGVQLGKCSVTRFSDGEVQINIEESIRGCDCY 60

Query: 61  IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVVPYFGYSRQDRIAASREPITAKLVAN 120
           IIQSTS PVN+++ ELLIM+DA KRASA T+NIV+PY+GY+RQDR A SREPITAKL AN
Sbjct: 61  IIQSTSDPVNEHIMELLIMVDALKRASAKTINIVIPYYGYARQDRKARSREPITAKLFAN 120

Query: 121 MLVKAGVDRVLTLDLHAVQVQGFFDIPVDNLFTVPLFAEHYNQLGLSGEDVVVVSPKNSG 180
           +L  AG  RV+ LDLHA Q+QGFFDIP+D+L  VP+  E++   G + ED+V+VSP + G
Sbjct: 121 LLETAGATRVIALDLHAPQIQGFFDIPIDHLMGVPILGEYFE--GKNLEDIVIVSPDHGG 178

Query: 181 IKRARSLAEYLDSPIAIIDYAQD-DSEREEGYIIGEVEGKKAIIIDDILNTGKTFAEAAK 239
           + RAR LA+ L +PIAIID  +   +  E   I+G +EGK AI+IDDI++T  T   AA
Sbjct: 179 VTRARKLADRLKAPIAIIDKRRPRPNVAEVMNIVGNIEGKTAILIDDIIDTAGTITLAAN 238

Query: 240 ILERGGATEIYAVASHGLFAGGAADILESAPIREIIVTDSV-LSKERIPSNIKYLTASHL 298
            L   GA E+YA  +H + +G A + + ++ I+E++VT+S+ L +E+     K L+   L
Sbjct: 239 ALVENGAKEVYACCTHPVLSGPAVERINNSTIKELVVTNSIKLPEEKKIERFKQLSVGPL 298

Query: 299 IADAIIRIHERKPLSPLFS 317
           +A+AIIR+HE++ +S LFS
Sbjct: 299 LAEAIIRVHEQQSVSYLFS 317
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5267> which encodes the amino acid sequence <SEQ ID 5268>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1830(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 278/324 (85%), Positives = 305/324 (93%)

Query: 1   MAEQYADKQIKLFSLTANREIAEKISQASGIPLGKMSSRQFSDGEIMINIEETVRGDDIY 60
           M E+YADKQIKLFSLT+N  IAEKI++A+GIPLGKMSSRQFS+GEIMINIEETVRGDDIY
Sbjct: 1   MTERYADKQIKLFSLTSNLPIAEKIAKAAGIPLGKMSSRQFSNGEIMINIEETVRGDDIY 60

Query: 61  IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVVPYFGYSRQDRIAASREPITAKLVAN 120
           IIQSTSFPVNDNLWELLIMIDACKRASANTVNIV+PYFGYSRQDR+A  REPITAKLVAN
Sbjct: 61  IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVLPYFGYSRQDRVAKPREPITAKLVAN 120

Query: 121 MLVKAGVDRVLTLDLHAVQVQGFFDIPVDNLFTVPLFAEHYNQLGLSGEDVVVVSPKNSG 180
           ML KAG+DRV+TLDLHAVQVQGFFDIPVDNLFTVPLFAE Y++LGLSG DVVVVSPKNSG
Sbjct: 121 MLTKAGIDRVVTLDLHAVQVQGFFDIPVDNLFTVPLFAERYSKLGLSGSDVVVVSPKNSG 180

Query: 181 IKRARSLAEYLDSPIAIIDYAQDDSEREEGYIIGEVEGKKAIIIDDILNTGKTFAEAAKI 240
           IKRARSLAEYLDSPIAIIDYAQDDSERE+GYIIG+V GKKAI+IDDILNTGKTFAEAAKI
Sbjct: 181 IKRARSLAEYLDSPIAIIDYAQDDSEREQGYIIGDVSGKKAILIDDILNTGKTFAEAAKI 240

Query: 241 LERGGATEIYAVASHGLFAGGAADILESAPIREIIVTDSVLSKERIPSNIKYLTASHLIA 300
           LER GAT+ YAVASHGLFAGGAAD+LE+API+EIIVTDSV +K R+P N+ YL+AS LIA
```

```
Sbjct: 241 LERSGATDTYAVASHGLFAGGAADVLETAPIKEIIVTDSVKTKNRVPENVTYLSASDLIA 300

Query: 301 DAIIRIHERKPLSPLFSYRSDKKD 324
           +AIIRIHER+PLSPLFSY+   K+
Sbjct: 301 EAIIRIHERRPLSPLFSYQPKGKN 324
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1697

A DNA sequence (GBSx1801) was identified in *S.agalactiae* <SEQ ID 5269> which encodes the amino acid sequence <SEQ ID 5270>. This protein is predicted to be Fe-S cluster formation protein. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1981(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04979 GB:AP001511 Fe-S cluster formation protein [Bacillus halodurans]
 Identities = 174/373 (46%), Positives = 237/373 (62%), Gaps = 6/373 (1%)

Query: 3   IYLDNAATTALTPSVIEKMTNVMTSNYGNPSSIHTFGRQANQLLRECRQIIAEYLNVNSR 62
           IYLD+AAT+ + P VI+ M       +GNPSSIH FGR+A Q + E R  IA  L   +
Sbjct: 4   IYLDHAATSPVHPEVIQAMLPYYEEQFGNPSSIHQFGRRARQGVDEARGTIARLLQADPS 63

Query: 63  EIIFTSGGTESNNTAIKGYALANQLKGKHIITSEIEHHSVLHTMTYLSERFGFDITYLKP 122
           E IFTSGGTE++N AI GYA  ++ KG HIITS++EHH+VLH    L E  GF++TY+
Sbjct: 64  EFIFTSGGTEADNLAIFGYAYQHRGKGNHIITSQVEHHAVLHACQEL-EHQGFEVTYVPV 122

Query: 123 NH-GQITAKDVQEALRDDTIMVSLMFVNNETGDFLPIQEIGQLLRNHQAVFHVDAVQVFS 181
           +  G+++ +DV++ALRDDTI+V+LM+ NNE G   PI EIG LL++HQAV H DAVQ F
Sbjct: 123 DQTGRVSVEDVRQALRDDTILVTLMYGNNEVGTIQPIAEIGALLQDHQAVLHTDAVQAFG 182

Query: 182 KMELDPHSLGIDFLAASAHKFHGPKGVGILYCAPH-HFDSLLHGGDQEEKRRASTENIIG 240
            + ++   L +D L+ SAHK +GPKGVG+LY          L+GG+QE K+RA TEN+
Sbjct: 183 AISIELDHLPVDMLSVSAHKINGPKGVGLLYVRDGIVLKPALYGGEQERKKRAGTENVAA 242

Query: 241 IAGMSQALTDATTNTLKNWTHISQLRTTFLDAISD--LDFYLNNGQDC-LPHVLNIGFPG 297
           I G ++A+  A  N  +          TF D      + F +N  Q   LPH+ N+ FPG
Sbjct: 243 IIGFAKAVEIAIANREERQKAYFDYCQTFFDQFQQEGVQFVMNGHQTWRLPHIFNVSFPG 302

Query: 298 QNNGLLLTQLDLAGFAVSTGSACTAGTVEPSHVLTSLYGANSPRLNESIRISFSELNTQE 357
            +   LL  LDLAG A S+GSACTAG++EPSHVL +++G++S  +   +R SF   NT+E
Sbjct: 303 VHVEALLVNLDLAGIAASSGSACTAGSIEPSHVLVAMHGSDSELVTSGVRFSFGLGNTKE 362

Query: 358 EILELAKTLRKII 370
            +   AK   KI+
Sbjct: 363 HVQWAAKETAKIV 375
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5271> which encodes the amino acid sequence <SEQ ID 5272>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1477(Affirmative) < succ>
```

```
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 235/370 (63%), Positives = 285/370 (76%)

Query: 2   MIYLDNAATTALTPSVIEKMTNVMTSNYGNPSSIHTFGRQANQLLRECRQIIAEYLNVNS 61
            M Y DNAATT L+P+VI  MT  M  N+GNPSSIH +GR+AN++LRECRQ IA  L  +
Sbjct: 1   MTYFDNAATTPLSPNVIRAMTAAMQDNFGNPSSIHFYGRRANKILRECRQAIARNLGASE 60

Query: 62  REIIFTSGGTESNNTAIKGYALANQLKGKHIITSEIEHHSVLHTMTYLSERFGFDITYLK 121
            ++II TSGGTESNN AIKGYALA+Q KGKH+IT+ IEHHSVLHTM YL ERFGF++TYL
Sbjct: 61  QQIIVTSGGTESNNMAIKGYALAHQAKGKHLITTTIEHHSVLHTMAYLEERFGFEVTYLP 120

Query: 122 PNHGQITAKDVQEALRDDTIMVSLMFVNNETGDFLPIQEIGQLLRNHQAVFHVDAVQVFS 181
              +GQI   D+++ALRDDTI+VS+M+ NNETGD LPI++IG LL++HQA FHVDAVQ
Sbjct: 121 CQNGQINLSDLKQALRDDTILVSIMYANNETGDLLPIKDIGNLLKDHQAAFHVDAVQAVG 180

Query: 182 KMELDPHSLGIDFLAASAHKFHGPKGVGILYCAPHHFDSLLHGGDQEEKRRASTENIIGI 241
            K+++ P  LGIDFL+ASAHKFHGPKG G LY      D LLHGGDQE KRRASTEN++GI
Sbjct: 181 KLKIIPSELGIDFLSASAHKFHGPKGCGFLYSNGQPIDPLLHGGDQEGKRRASTENMLGI 240

Query: 242 AGMSQALTDATTNTLKNWTHISQLRTTFLDAISDLDFYLNNGQDCLPHVLNIGFPGQNNG 301
             GM+QALTDA T   ++  HI  LR   +  +  L +Y+N G   LPHVLNIGF G  N
Sbjct: 241 IGMAQALTDAMTCLDQSTDHIISLRHHLISLLEGLPYYINQGTHYLPHVLNIGFLGYQNT 300

Query: 302 LLLTQLDLAGFAVSTGSACTAGTVEPSHVLTSLYGANSPRLNESIRISFSELNTQEEILE 361
            +LLTQLDLAG AVSTGSACTAG V PSHVL + YG +S RL ESIRISFS+ N+ E++ +
Sbjct: 301 ILLTQLDLAGIAVSTGSACTAGAVNPSHVLAAYYGDDSSRLKESIRISFSDQNSIEDVNQ 360

Query: 362 LAKTLRKIIG 371
            LA+TL+ I+G
Sbjct: 361 LAQTLKNILG 370
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1698

A DNA sequence (GBSx1802) was identified in *S.agalactiae* <SEQ ID 5273> which encodes the amino acid sequence <SEQ ID 5274>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2753(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12416 GB:Z99107 ydiH [Bacillus subtilis]
 Identities = 96/202 (47%), Positives = 140/202 (68%), Gaps = 4/202 (1%)

Query: 7   IPKATAKRLSLYYRIFKRFNTDGIEKASSKQIADALGIDSATVRRDFSYFGELGRRGFGY 66
            IP+ATAKRL LYYR  K  +  G ++ SS +++DA+ +DSAT+RRDFSYFG LG++G+GY
Sbjct: 8   IPQATAKRLPLYYRFLKNLHASGKQRVSSAELSDAVKVDSATIRRDFSYFGALGKKGYGY 67

Query: 67  DVKKLMNFFAEILNDHSTTNVMLVGCGNIGRALLHYRFHDRNKMQISMAFDLDSNDLVGK 126
            +V  L++FF + L+    T+V+L+G GN+G A LHY F   N  +ISMAFD++ + +
Sbjct: 68  NVDYLLSFFRKTLDQDEMTDVILIGVGNLGTAFLHYNFTKNNNTKISMAFDINESKI--G 125

Query: 127 TTEDGIPVYGISTINDHLIDSDIETAILTVPSTEAQEVADILVKAGIKGILSFSPVHLTL 186
            T   G+PVY +  +  H+ D  +  AILTVP+  AQ + D LV  GIKGIL+F+P  L +
```

```
Sbjct: 126 TEVGGVPVYNLDDLEQHVKDESV--AILTVPAVAAQSITDRLVALGIKGILNFTPARLNV 183

Query: 187 PKDIIVQYVDLTSELQTLLYFM 208
           P+ I + ++DL  ELQ+L+YF+
Sbjct: 184 PEHIRIHHIDLAVELQSLVYFL 205
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5275> which encodes the amino acid sequence <SEQ ID 5276>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2313(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 167/210 (79%), Positives = 189/210 (89%)

Query: 1   MIMDKSIPKATAKRLSLYYRIFKRFNTDGIEKASSKQIADALGIDSATVRRDFSYFGELG 60
           +++DKSIPKATAKRLSLYYRIFKRF+ D +EKASSKQIADA+GIDSATVRRDFSYFGELG
Sbjct: 1   VVIDKSIPKATAKRLSLYYRIFKRFHADQVEKASSKQIADAMGIDSATVRRDFSYFGELG 60

Query: 61  RRGFGYDVKKLMNFFAEIILNDHSTTNVMLVGCGNIGRALLHYRFHDRNKMQISMAFDLDS 120
           RRGFGYDV KLMNFFA++LNDHSTTNV+LVGCGNIGRALLHYRFHDRNKMQI+M FD D
Sbjct: 61  RRGFGYDVTKLMNFFADLLNDHSTTNVILVGCGNIGRALLHYRFHDRNKMQIAMGFDTDD 120

Query: 121 NDLVGKTTEDGIPVYGISTINDHLIDSDIETAILTVPSTEAQEVADILVKAGIKGILSFS 180
           N LVG  T D IPV+GIS++ + + ++DIETAILTVPS  AQEV D L++AGIKGILSF+
Sbjct: 121 NALVGTKTADNIPVHGISSVKERIANTDIETAILTVPSIHAQEVTDQLIEAGIKGILSFA 180

Query: 181 PVHLTLPKDIIVQYVDLTSELQTLLYFMNQ 210
           PVHL +PK +IVQ VDLTSELQTLLYFMNQ
Sbjct: 181 PVHLQVPKGVIVQSVDLTSELQTLLYFMNQ 210
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1699

A DNA sequence (GBSx1803) was identified in *S.agalactiae* <SEQ ID 5277> which encodes the amino acid sequence <SEQ ID 5278>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2966(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9847> which encodes amino acid sequence <SEQ ID 9848> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14764 GB:Z99118 similar to DNA repair protein [Bacillus subtilis]
 Identities = 90/210 (42%), Positives = 136/210 (63%)

Query: 24  PRERLVDLGADRLSNQELLAILLRTGIKEKPVLEISTQILENISSLADFGQLSLQELQSI 83
           PRERL+ +GA+ L+N ELLAILLRTG K + VL++S ++L +   L    + S++EL SI
Sbjct: 19  PRERLLKVGAENLANHELLAILLRTGTKHESVLDLSNRLLRSFDGLRLLKEASVEELSSI 78
```

```
Query: 84  KGIGQVKSVEIKAMLELAKRIHKAEYDRKEQILSSEQLARKMMLELGDKKQEHLVAIYMD 143
            GIG VK+++I A +EL  RIHK   +    I S E  A  +M ++    QEH V +Y++
Sbjct: 79  PGIGMVKAIQILAAVELGSRIHKLANEEHFVIRSPEDGANLVMEDMRFLTQEHFVCLYLN 138

Query: 144 TQNRIIEQRTIFIGTVRRSVAEPREILHYACKNMATSLIIIHNHPSGSPKPSESDLSFTK 203
           T+N++I +RT+FIG++  S+  PRE+   A K  A S I +HNHPSG P PS  D+  T+
Sbjct: 139 TKNQVIHKRTVFIGSLNSSIVHPREVFKEAFKRSAASFICVHNHPSGDPTPSREDIEVTR 198

Query: 204 KIKRSCDHLGIVCLDHIIVGKNKYYSFREE 233
           ++    + +GI  LDH+++G  K+ S +E+
Sbjct: 199 RLFECGNLIGIELLDHLVIGDKKFVSLKEK 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5279> which encodes the amino acid sequence <SEQ ID 5280>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 145/225 (64%), Positives = 182/225 (80%)

Query: 12  MYHIELKKEALLPRERLVDLGADRLSNQELLAILLRTGIKEKPVLEISTQILENISSLAD 71
           MY I+      +PRERL+ LGA+ LSNQELLAILLRTG KEK VLE+S+ +L ++ SLAD
Sbjct: 1   MYSIKCDDNKAMPRERLMRLGAESLSNQELLAILLRTGNKEKHVLELSSYLLSHLDSLAD 60

Query: 72  FGQLSLQELQSIKGIGQVKSVEIKAMLELAKRIHKAEYDRKEQILSSEQLARKMMLELGD 131
           F ++SLQELQ + GIG+VK++EIKAM+EL  RI   +    + +L+S Q+A KMM  LGD
Sbjct: 61  FKKMSLQELQHLAGIGKVKAIEIKAMIELVSRILATDKTLTDSVLTSVQVAEKMMAALGD 120

Query: 132 KKQEHLVAIYMDTQNRIIEQRTIFIGTVRRSVAEPREILHYACKNMATSLIIIHNHPSGS 191
           KKQEHLV +Y+D QNRIIE++TIFIGTVRRS+AEPREIL+YACKNMATSLI+IHNHPSG+
Sbjct: 121 KKQEHLVVLYLDNQNRIIEEKTIFIGTVRRSLAEPREILYYACKNMATSLIVIHNHPSGN 180

Query: 192 PKPSESDLSFTKKIKRSCDHLGIVCLDHIIVGKNKYYSFREEADI 236
            +PS +D  FT+KIKRSC+ LGI+CLDHIIV    YYSFRE++ +
Sbjct: 181 IEPSSNDYCFTEKIKRSCEDLGIICLDHIIVSYKDYYSFREKSTL 225
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1700

A DNA sequence (GBSx1804) was identified in *S.agalactiae* <SEQ ID 5281> which encodes the amino acid sequence <SEQ ID 5282>. This protein is predicted to be a permease. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.86   Transmembrane  258 - 274 ( 255 - 290)
   INTEGRAL    Likelihood = -7.32   Transmembrane   89 - 105 (  79 - 109)
   INTEGRAL    Likelihood = -4.88   Transmembrane  176 - 192 ( 170 - 194)
   INTEGRAL    Likelihood = -4.78   Transmembrane  339 - 355 ( 326 - 359)
   INTEGRAL    Likelihood = -4.57   Transmembrane  237 - 253 ( 236 - 257)
   INTEGRAL    Likelihood = -3.98   Transmembrane   39 -  55 (  38 -  59)
   INTEGRAL    Likelihood = -3.40   Transmembrane  292 - 308 ( 282 - 308)
   INTEGRAL    Likelihood = -1.38   Transmembrane  317 - 333 ( 317 - 333)
   INTEGRAL    Likelihood = -0.27   Transmembrane    8 -  24 (   8 -  24)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05771 GB:AF051356 putative permease [Streptococcus mutans]
 Identities = 88/366 (24%), Positives = 175/366 (47%), Gaps = 27/366 (7%)

Query: 3   FEKRQVYYVVITFAICYAIQAYW---GAVSNILTTLHKAIF-PFLMGAGIAYIINIVMSV 58
           F+  ++++  +   +   I   W   G++ N   ++ K  F PFL+G  + YI N +++
Sbjct: 2   FKSSKLFFWTVEILLVTLILFIWRQMGSIFNPFFSVAKTFFLPFLLGGFLYYITNPIVTF 61

Query: 59  YERLYIKLFKGSRLLMAIKRSVSMILSYATFIGLIVWLFSIVIPDLISSLSSLLVIDTGA 118
            E  +            IKR   + L +A  + L+V+  + +IP+LI+ L+ L+
Sbjct: 62  LENRF-----------KIKRIWGITLIFAVLLSLLVFSITSLIPNLINQLTDLISASQNI 110

Query: 119 LAKLVNNLNENKQISEVLNYMGTDKDLVSTLSGYSQQILKQVLSVLTNLLTSVSSIAATL 178
              L +  NE K      N      D+   L  ++   +  + +VL ++  SVSSI   +
Sbjct: 111 YVGLQDLFNEWKSNPAFKNI-----DIPVLLKQFNLSYVDILTNVLDSVTVSVSSIVYMI 165

Query: 179 LNVFVSFIFS----IYVLANKEQLGRQFNLLIDTYLGSTGKTFHYVRHILHQRFHGFFVS 234
            N  +  + +     Y+L +K+ L     +L  T L +       + + +++    +
Sbjct: 166 TNTVMILVLTPVILFYLLKDKDGL---MPMLDRTILKNDRHNISQLLNQMNKTISRYISG 222

Query: 235 QTLEAMILGSLTVIGMLIFQFPYALTVGVLVAFTALIPVVGAYIGVTIGFILIATESLTE 294
             ++A  +    +IG  I    YA    ++   T +IP VG Y+G+T   +        +
Sbjct: 223 VAIDAAFIFVFALIGYQIMGVQYAFLFALVAGITNVIPYVGPYLGLTPVVLAYVVSDPKK 282

Query: 295 AFLFVLFLILLQQFEGNVIYPKVVGGSIGLPSMWVLMAITIGGALWGILGMLLAVPVAAT 354
             + +++++ LQQ +GN++YP+VVG ++ +  + +++ + +GG + G++GML+AVP  A
Sbjct: 283 MIIAIIYIMTLQQIDGNIVYPRVVGSTMKIHPLTIMVLLVLGGNIAGLVGMLVAVPAYAI 342

Query: 355 IYQIVK 360
           I +IVK
Sbjct: 343 IKEIVK 348
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5283> which encodes the amino acid sequence <SEQ ID 5284>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70    Transmembrane   87 - 103 (  83 - 116)
   INTEGRAL    Likelihood = -7.27    Transmembrane  178 - 194 ( 166 - 202)
   INTEGRAL    Likelihood = -6.74    Transmembrane  278 - 294 ( 256 - 297)
   INTEGRAL    Likelihood = -5.41    Transmembrane  299 - 315 ( 295 - 321)
   INTEGRAL    Likelihood = -4.46    Transmembrane   14 -  30 (  13 -  32)
   INTEGRAL    Likelihood = -3.56    Transmembrane  340 - 356 ( 333 - 366)
   INTEGRAL    Likelihood = -3.35    Transmembrane  258 - 274 ( 256 - 277)

----- Final Results -----
              bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC05771 GB:AF051356 putative permease [Streptococcus mutans]
 Identities = 87/373 (23%), Positives = 168/373 (44%), Gaps = 41/373 (10%)

Query: 10  FEKKQVFYLVLTFILCYGILANWRNGTAIVTTIYKTS----LPFFYGAAGAYIVNIVMSA 65
           F+  ++F+  +  +L   IL  WR   +I    +  +    LPF  G    YI N +++
Sbjct: 2   FKSSKLFFWTVEILLVTLILFIWRQMGSIFNPFFSVAKTFFLPFLLGGFLYYITNPIVTF 61

Query: 66  YEKVYVYIFKDWSHVLKVKRGICLLLAYLTFFILITWIISIVIPDLITSISTLTKFDT-- 123
```

```
            E  +           K+KR   + L +     L+ + I+ +IP+LI  ++ L
Sbjct: 62  LENRF-----------KIKRIWGITLIFAVLLSLLVFSITSLIPNLINQLTDLISASQNI 110

Query: 124 -ITIQEVVNNLEHNKLLARTIQYIGGDGKLTETIANYSQQLLKQFLTVLTNILTSVTVIA 182
            + +Q++ N  + N                   I    +Q    ++ +LTN+L SVTV
Sbjct: 111 YVGLQDLFNEWKSNPAFKNI------------DIPVLLKQFNLSYVDILTNVLDSVTVSV 158

Query: 183 SAIINLFISFVFSL--------YVLASKEDLCRQGNTLVDTYTGKYAKRIHYLLELLHQR 234
           S+I+ +  + V  L        Y+L  K+ L      L  T        I  LL  +++
Sbjct: 159 SSIVYMITNTVMILVLTPVILFYLLKDKDGLMPM---LDRTILKNDRHNISQLLNQMNKT 215

Query: 235 FHGFFVSQTLEAMILGSLTASGMFILRLPFAGTIGVLVAFTALIPVIGASIGAAIGFILI 294
              +     ++A  +      G  I+ + +A     ++  T +IP +G  +G     +
Sbjct: 216 ISRYISGVAIDAAFIFVFALIGYQIMGVQYAFLFALVAGITNVIPYVGPYLGLTPVVLAY 275

Query: 295 MTQSMSQAIIFIIFLIILQQIEGNFIYPKVVGGSIGLPAMWVLMAITIGASLKGIVGMII 354
           +     + II II+++ LQQI+GN +YP+VVG ++ +  + +++ + +G ++ G+VGM++
Sbjct: 276 VVSDPKKMIIAIIYIMTLQQIDGNIVYPRVVGSTMKIHPLTIMVLLVLGGNIAGLVGMLV 335

Query: 355 AVPLAATLYQVIK 367
           AVP  A + +++K
Sbjct: 336 AVPAYAIIKEIVK 348
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 218/370 (58%), Positives = 291/370 (77%)

Query: 1   MKFEKRQVYYVVITFAICYAIQAYWGAVSNILTTLHKAIFPFLMGAGIAYIINIVMSVYE 60
           MKFEK+QV+Y+V+TF +CY I A W   + I+TT++K   PF  GA  AYI+NIVMS YE
Sbjct: 8   MKFEKKQVFYLVLTFILCYGILANWRNGTAIVTTIYKTSLPFFYGAAGAYIVNIVMSAYE 67

Query: 61  RLYIKLFKGSRLLMAIKRSVSMILSYATFIGLIVWLFSIVIPDLISSLSSLLVIDTGALA 120
           ++Y+ +FK    ++ +KR + ++L+Y TF  LI W+ SIVIPDLI+S+S+L   DT  +
Sbjct: 68  KVYVYIFKDWSHVLKVKRGICLLLAYLTFFILITWIISIVIPDLITSISTLTKFDTITIQ 127

Query: 121 KLVNNLNENKQISEVLNYMGTDKDLVSTLSGYSQQILKQVLSVLTNLLTSVSSIAATLLN 180
           ++VNNL  NK ++  + Y+G D  L  T++ YSQQ+LKQ L+VLTN+LTSV+ IA+ ++N
Sbjct: 128 EVVNNLEHNKLLARTIQYIGGDGKLTETIANYSQQLLKQFLTVLTNILTSVTVIASAIIN 187

Query: 181 VFVSFIFSIYVLANKEQLGRQFNLLIDTYLGSTGKTFHYVRHILHQRFHGFFVSQTLEAM 240
           +F+SF+FS+YVLA+KE L RQ N L+DTY G   K  HY+  +LHQRFHGFFVSQTLEAM
Sbjct: 188 LFISFVFSLYVLASKEDLCRQGNTLVDTYTGKYAKRIHYLLELLHQRFHGFFVSQTLEAM 247

Query: 241 ILGSLTVIGMLIFQFPYALTVGVLVAFTALIPVVGAYIGVTIGFILIATESLTEAFLFVL 300
           ILGSLT  GM I + P+A T+GVLVAFTALIPV+GA IG  IGFILI T+S+++A +F++
Sbjct: 248 ILGSLTASGMFILRLPFAGTIGVLVAFTALIPVIGASIGAAIGFILIMTQSMSQAIIFII 307

Query: 301 FLILLQQFEGNVIYPKVVGGSIGLPSMWVLMAITIGGALWGILGMLLAVPVAATIYQIVK 360
           FLI+LQQ EGN IYPKVVGGSIGLP+MWVLMAITIG +L GI+GM++AVP+AAT+YQ++K
Sbjct: 308 FLIILQQIEGNFIYPKVVGGSIGLPAMWVLMAITIGASLKGIVGMIIAVPLAATLYQVIK 367

Query: 361 DHIIKRQTLR 370
           D+I KRQ ++
Sbjct: 368 DNIQKRQAIQ 377
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1701

A DNA sequence (GBSx1805) was identified in *S.agalactiae* <SEQ ID 5285> which encodes the amino acid sequence <SEQ ID 5286>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1081(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9849> which encodes amino acid sequence <SEQ ID 9850> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA69226 GB:U29579 6-phospho-beta-glucosidase [Escherichia coli]
 Identities = 290/478 (60%), Positives = 369/478 (76%), Gaps = 2/478 (0%)

Query: 2   MVKQVFPKGFLWGGATAANQCEGAYNVDGRGLANVDVVPTGEDRFAIISGQKKMFDFEEG 61
           M   VFP+ FLWGGA AANQ EGA+    +GL  VD++P GE R A+  G +K F   +
Sbjct: 1   MKMSVFPESFLWGGALAANQSEGAFREGDKGLTTVDMIPHGEHRMAVKLGLEKRFQLRDD 60

Query: 62  YFYPAKESIDFYHHYKEDLALLAEMGFKTYRMSIAWTRIFPKGDELYPNEAGLQFYENIF 121
            FYP+ E+ DFYH YKED+AL+AEMGFK +R SIAW+R+FP+GDE+ PN+ G+ FY ++F
Sbjct: 61  EFYPSHEATDFYHRYKEDIALMAEMGFKVFRTSIAWSRLFPQGDEITPNQQGIAFYRSVF 120

Query: 122 KECRKYGIEPLVTITHFDCPIYLIKHYGGWRSRKMIGFYERLVRALFTRFKGLVKYWLTF 181
           +EC+KYGIEPLVT+ HFD P++L+  YG WR+RK++ F+ R  R  F  F GLVKYWLTF
Sbjct: 121 EECKKYGIEPLVTLCHFDVPMHLVTEYGSWRNRKLVEFFSRYARTCFEAFDGLVKYWLTF 180

Query: 182 NEINMILHAPFMGAGLYFEDGENQEQIKYQAAHHELVASAIAVKIAHEVDPNNQIGCMLA 241
           NEIN++LH+PF GAGL FE+GENQ+Q+KYQAAHH+LVASA+A KIAHEV+P NQ+GCMLA
Sbjct: 181 NEINIMLHSPFSGAGLVFEEGENQDQVKYQAAHHQLVASALATKIAHEVNPQNQVGCMLA 240

Query: 242 AGQYYPNTCHPQDYWASMQKNRENYFFIDVQARGKYPNYAKKHFEHLGISIQMTAEDLAL 301
            G +YP +C P+D WA+++K+REN FFIDVQARG YP Y+ + F   G++I     D  +
Sbjct: 241 GGNFYPYSCKPEDVWAALEKDRENLFFIDVQARGTYPAYSARVFREKGVTINKAPGDDEI 300

Query: 302 LRDYTVDFISFSYYSSRVASGNPTVSEQVQENIFASLKNPYLKSSEWGWQIDPLGLRITL 361
           L++ TVDF+SFSYY+SR AS     +     N+  SL+NPYL+ S+WGW IDPLGLRIT+
Sbjct: 301 LKN-TVDFVSFSYYASRCASAEMNANNSSAANVVKSLRNPYLQVSDWGWGIDPLGLRITM 359

Query: 362 NAIWDRYQKPMFIVENGLGAVDIPDENGYVEDDYRIDYLRQHIAAMRDAIYVDGVNLIGY 421
           N ++DRYQKP+F+VENGLGA D    NG + DDYRI YLR+HI AM +AI  DG+ L+GY
Sbjct: 360 NMMYDRYQKPLFLVENGLGAKDEFAANGEINDDYRISYLREHIRAMGEAI-ADGIPLMGY 418

Query: 422 TTWGCIDLVSAGTGEMEKRYGFIYVDRNNKGEGTLKRYKKKSFYWYKKVIASNGSQIE 479
           TTWGCIDLVSA TGEM KRYGF++VDR++ G GTL R +KKSF+WYKKVIASNG  +E
Sbjct: 419 TTWGCIDLVSASTGEMSKRYGFVFVDRDDAGNGTLTRTRKKSFWWYKKVIASNGEDLE 476
```

There is also homology to SEQ ID 5288.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1702

A DNA sequence (GBSx1806) was identified in *S.agalactiae* <SEQ ID 5289> which encodes the amino acid sequence <SEQ ID 5290>. This protein is predicted to be platelet-activating factor acetylhydrolase isoform Ib beta subunit, pu. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5323(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC27974 GB:AF016048 platelet-activating factor acetylhydrolase
           alpha 2 subunit [Rattus norvegicus]
 Identities = 43/177 (24%), Positives = 84/177 (47%), Gaps = 9/177 (5%)

Query: 28  QEGAIVFTGDSIVEF---FPLKKHLGRDYPLVNRGVAGSDTYWLLENLRTQVWELLPSKV 84
            +E  ++F GDS+V+    + + + L      +N G+ G  T  +L  L+    E +  KV
Sbjct: 38  KEPDVLFVGDSMVQLMQQYEIWRELFSPLHALNFGIGGDTTRHVLWRLKNGELENIKPKV 97

Query: 85  FIL-IGTNDIGLGHSQSEIIANITDIIAEIRAESYMTEINILSVLPVSEEDDYIERVKVR 143
             ++ +GTN+    ++  E+   I  I+  I       +I +L +LP  E+ + + +   +
Sbjct: 98  IVVWVGTNNHE--NTAEEVAGGIEAIVQLINTRQPQAKIIVLGLLPRGEKPNPLRQKNAK 155

Query: 144 NNQTIKALNKTLSVISGINYIELYDLLVDEKGQLASSFTKDGLHLTDQAYAKISETI 200
             NQ +K    +L  ++ +  +++    V   G ++     D LHLT   YAKI + +
Sbjct: 156 VNQLLKV---SLPKLANVQLLDIDGGFVHSDGAISCHDMFDFLHLTGGGYAKICKPL 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5291> which encodes the amino acid sequence <SEQ ID 5292>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5979(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 92/204 (45%), Positives = 133/204 (65%)

Query: 1   MLEVIDKALRDYQMKREQFFEINNQTVQEGAIVFTGDSIVEFFPLKKHLGRDYPLVNRGV 60
            MLE++ + LR YQ ++   +    NQ   +G IVF GDS++EFFPLKK  G   P++NRG+
Sbjct: 1   MLEIVSEELRHYQEQKLIEYRNKNQLAPKGGIVFAGDSLIEFFPLKKAFGSCLPIINRGI 60

Query: 61  AGSDTYWLLENLRTQVWELLPSKVFILIGTNDIGLGHSQSEIIANITDIIAEIRAESYMT 120
            AG D+ WLL  +  Q+ +L P  +F+LIG NDIGLG+ + I+  I ++I++IR+    +
Sbjct: 61  AGIDSQWLLRHFSVQITDLEPKHIFLLIGCNDIGLGYDKCHIVKTIVELISQIRSHCVYS 120

Query: 121 EINILSVLPVSEEDDYIERVKVRNNQTIKALNKTLSVISGINYIELYDLLVDEKGQLASS 180
            +I +LS+LPVS    Y + VK+R N  I A+NK L++I  + +I L   L DEKG L+
Sbjct: 121 QIYLLSLLPVSNNPRYQKTVKIRTNAMIDAINKDLAMIPTVEFINLNTCLKDEKGGLSDE 180

Query: 181 FTKDGLHLTDQAYAKISETIKLYL 204
             T DGLHL   AYAK++E IK Y+
Sbjct: 181 NTLDGLHLNFPAYAKLAEIIKSYI 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1703

A DNA sequence (GBSx1807) was identified in *S.agalactiae* <SEQ ID 5293> which encodes the amino acid sequence <SEQ ID 5294>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5226(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9851> which encodes amino acid sequence <SEQ ID 9852> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
           intergenic region. [Escherichia coli]
 Identities = 104/265 (39%), Positives = 154/265 (57%), Gaps = 4/265 (1%)

Query: 2   IKLIATDMDGTFLRSDKTYDKARFSSLLTLMEKYDIKFVAASGNLYDQLLLNFLEYPNRI 61
           IKLIA DMDGTFL   KTY++ RF +    M+   I+FV ASGN Y QL+  F E  N I
Sbjct: 4   IKLIAVDMDGTFLSDQKTYNRERFMAQYQQMKAQGIRFVVASGNQYYQLISFFPEIANEI 63

Query: 62  AYVAENGGRVIDQDGTLLKETYLSNDTVAAVLSYLYQNYPETLISLSGEKRSYLERRTPI 121
           A+VAENGG V+ + G  +    LS D  A V+ +L    PE  I   G+  +Y  ++
Sbjct: 64  AFVAENGGWVVSE-GKDVFNGELSKDAFATVVEHLLTR-PEVEIIACGKNSAYTLKKYDD 121

Query: 122 NRRTELEYYMPNFIYKDHLLPLDDDRYFQMTLWVNENLVSEMLLDISEHFKNHHIRLTSS 181
             +T  E Y     Y D+   L+D  +F+  L +++ L+ ++   + E   +  + +  +
Sbjct: 122 AMKTVAEMYYHRLEYVDNFDNLEDI-FFKFGLNLSDELIPQVQKALHEAIGDIMVSV-HT 179

Query: 182 GFGCIDVLPADVNKADGIAILLEKWGLKQDQVMVFGDGGNDVEMLRAANISYAMSNAPEE 241
           G G ID++   V+KA+G+  L + WG+   +V+VFGDGGND+EMLR A  S+AM NA
Sbjct: 180 GNGSIDLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSA 239

Query: 242 IKAIAKYQTVSNDQDGVLETIENFL 266
           + A AKY+  SN+++GVL+ I+  L
Sbjct: 240 VVAAAKYRAGSNNREGVLDVIDKVL 264
```

There is also homology to SEQ ID 1158.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1704

A DNA sequence (GBSx1808) was identified in *S.agalactiae* <SEQ ID 5295> which encodes the amino acid sequence <SEQ ID 5296>. This protein is predicted to be transcriptional regulator (AraC/XylSfamily). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF89977 GB:AF206272 transcriptional regulator [Streptococcus mutans]
 Identities = 195/287 (67%), Positives = 237/287 (81%)

Query: 5   DNLLSHNLEDNRHLLPYEHMHTEVRNGYPDILFHWHPELEISYVHEGTARYHIDYDFFNS 64
           D    H +  +  LLPY+   T + NGYPD LFHWHPELEISY++EGTA+YHIDYD+FNS
Sbjct: 10  DENFKHEINFDNDLLPYKIYQTTIANGYPDTLFHWHPELEISYIYEGTAQYHIDYDYFNS 69

Query: 65  QSGDIILIRPNGMHSIHPIENKEHITDSIKFHLDLIGYSIVDQVSLRYLQPLQTSSFKFI 124
           Q+ DIIL+RPNGMHSIHPI+NK     ++ FHLDL+GYS++DQ+SLRYLQPLQ S+FK +
Sbjct: 70  QTDDIILVRPNGMHSIHPIKNKMQKAQTLLFHLDLVGYSLLDQISLRYLQPLQNSTFKLV 129

Query: 125 QCIKPSMTGYNDIKNCLFDIFNISKEENRHFELLLKAKLNELLYLLYYHQYVIKKHTDDT 184
            CIKP M GY DIKNCLF IF+I + + RHFELLLKAKL EL+YLLY+HQYV++KH+DD
Sbjct: 130 PCIKPDMLGYQDIKNCLFAIFDIYQRQGRHFELLLKAKLQELIYLLYFHQYVLRKHSDDM 189
```

```
Query: 185 YRKNERIRDLIDYINNNYQQNLTIEFLADYMGYSKTHFMTVFKQHTGTSCTEFIIQVRLN 244
           YRKNE+IR+LIDYI+ +YQ+ L+I  LAD +GYSKTHFMTVFKQHTGTSCT+FIIQ RL+
Sbjct: 190 YRKNEKIRELIDYIHQHYQEKLSIISLADIIGYSKTHFMTVFKQHTGTSCTDFIIQFRLS 249

Query: 245 KASEHLINSTTAIIDIANSVGFNNLSNFNRQFKRYYHTTPRQYRKQF 291
           KA + L+NS   I+++A+ VGF NLSNFNRQFKRYY  TP QYRKQF
Sbjct: 250 KACDLLVNSIKPILEVASEVGFTNLSNFNRQFKRYYQITPSQYRKQF 296
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5297> which encodes the amino acid sequence <SEQ ID 5298>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 43/169 (25%), Positives = 83/169 (48%), Gaps = 16/169 (9%)

Query: 136 DIKNCLFDIFNISKEENRHFELLLKAKLNELLYLLYYHQYV------IKKHTDDTYRKN- 188
           D+K+  F +F+   +  R F +L K     ++  ++  Q +      +KK  D T + N
Sbjct: 319 DVKHVSFLLFS---DIYRQFPILDKMTYLSMVKTIHDSQSIDCILRELKKVLDVTNQNNS 375

Query: 189 ------ERIRDLIDYINNNYQQNLTIEFLADYMGYSKTHFMTVFKQHTGTSCTEFIIQVR 242
                 + + + ID I   Y Q LT++ +AD +  +  +    FK  T  S T+++  VR
Sbjct: 376 PEKRYSDLVSETIDCIRKEYHQELTLKAIADRLHVNGVYLGQCFKNETERSFTQYLNHVR 435

Query: 243 LNKASEHLINSTTAIIDIANSVGFNNLSNFNRQFKRYYHTTPRQYRKQF 291
           + KA + L+ +  +I +IA   G+N    F + FK+    +P+++R ++
Sbjct: 436 IQKAQQLLLYTNQSINEIAYETGYNTNHYFIKMFKKLNGLSPKEFRDRY 484
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1705

A DNA sequence (GBSx1809) was identified in *S.agalactiae* <SEQ ID 5299> which encodes the amino acid sequence <SEQ ID 5300>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3705(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1706

A DNA sequence (GBSx1810) was identified in *S.agalactiae* <SEQ ID 5301> which encodes the amino acid sequence <SEQ ID 5302>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25   Transmembrane   59 -  75 (  56 -  82)
   INTEGRAL    Likelihood = -7.48   Transmembrane   23 -  39 (  12 -  41)
   INTEGRAL    Likelihood = -6.64   Transmembrane  231 - 247 ( 225 - 255)
   INTEGRAL    Likelihood = -5.15   Transmembrane  335 - 351 ( 333 - 355)
   INTEGRAL    Likelihood = -4.19   Transmembrane  309 - 325 ( 305 - 327)
   INTEGRAL    Likelihood = -4.14   Transmembrane  272 - 288 ( 268 - 292)
   INTEGRAL    Likelihood = -4.04   Transmembrane  402 - 418 ( 400 - 419)
   INTEGRAL    Likelihood = -3.88   Transmembrane  191 - 207 ( 190 - 208)
   INTEGRAL    Likelihood = -2.71   Transmembrane  365 - 381 ( 364 - 381)
   INTEGRAL    Likelihood = -1.86   Transmembrane  165 - 181 ( 164 - 182)

----- Final Results -----
               bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF96429 GB:AE004383 conserved hypothetical protein [Vibrio cholerae]
 Identities = 142/443 (32%), Positives = 241/443 (54%), Gaps = 20/443 (4%)

Query: 6   NEFQFSLESILGFVWRGIVVGLIAGFVVSIFRLAIEKIFLVVMELYKS--AHYQPIILLS 63
           N+F    ++    ++  ++VG++AG V + F  A+  +     +  KS    + P+ L +
Sbjct: 21  NQFLSKDKTPFSVLFLSLLVGILAGLVGTYFEQAVHLVSETRTDWLKSEIGSFLPLWLAA 80

Query: 64  ITVTSIIAAVIIGFFI--KSDPDIKGSGIPHVEGELKGMLSPDWFSIVWKKFIAGILAIS 121
             +++ +A   IG+F+  +  P+  GSGIP +EG + GM    W+ ++  KF  G+ A+
Sbjct: 81  FLISAFLA--FIGYFLVHRFAPEAAGSGIPEIEGAMDGMRPVRWWRVLPVKFFGGMGALG 138

Query: 122 SGLMLGREGPSIQLGAMTGKGIAQYLNASRMEKR-VLIASGAAAGLSAAFNAPIAGLLFV 180
           SG++LGREGP++Q+G   G+ I+        + R  L+A+GAA GL+AAFNAP+AG++FV
Sbjct: 139 SGMVLGREGPTVQMGGAVGRMISDIFRVKNEDTRHSLLAAGAAGGLAAAFNAPLAGIMFV 198

Query: 181 VEEIYHHFS-RLVWITALVASLV-ANFVSLNIFGLTPVLALPSELPSLNLNFYWIFLLMG 238
           +EE+   F   L+ + A++ S V AN V   I G   V+ +P +  +  L+   +FLL+G
Sbjct: 199 IEEMRPQFRYTLISVRAVIISAVAANIVFRVINGQDAVITMP-QYDAPELSTLGLFLLLG 257

Query: 239 LFLGILGFIYEWVIL----RFHVIYDYLGKLFHLPSHLYGILAVIFILPIGYYFPQLLGG 294
              G+ G ++ ++I      F   +    K + L   + G    + +L    Y P+L GG
Sbjct: 258 ALFGVFGVLFNYLITLAQDLFVKFHRNDRKRYLLTGSMIGGCFGLLLL----YVPELTGG 313

Query: 295 GNGLIVSLPRSNLSLMMLGLFFLIRFLWSMLSYSSGLPGGIFLPILALGSLAG-AFFAVG 353
           G  LI ++        +L L F+ R   ++L + SG PGGIF P+LALG+L G AF  +
Sbjct: 314 GISLIPTITNGGYGAGILLLLFVGRIFTTLLCFGSGAPGGIFAPMLALGTLFGYAFGLIA 373

Query: 354 MQYFGIISHQQISLFVVLGMAGYFGAISKAPLTAMILVTEMVGDLKQLMAIGIVTMVSYI 413
             +F  ++ +   +F + GM   F A  +AP+T ++LV EM  +   ++ + I ++ + I
Sbjct: 374 KMWFPELNIEP-GMFAIAGMGALFAATVRAPITGILLVIEMTNNYHLILPLIITSLGAVI 432

Query: 414 VMDLLKGEPIYEAMLAKMTFNPK 436
              LL G+PIY  +L +   N K
Sbjct: 433 FAQLLGGQPIYSQLLHRTLKNQK 455
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5303> which encodes the amino acid sequence <SEQ ID 5304>. Analysis of this protein sequence reveals the following:

```
    Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.68   Transmembrane   71 -  87 (  66 -  95)
   INTEGRAL    Likelihood = -9.45   Transmembrane   36 -  52 (  26 -  56)
```

```
INTEGRAL    Likelihood = -5.63    Transmembrane  346 - 362 ( 342 - 367)
INTEGRAL    Likelihood = -5.36    Transmembrane  376 - 392 ( 375 - 393)
INTEGRAL    Likelihood = -5.15    Transmembrane  413 - 429 ( 410 - 432)
INTEGRAL    Likelihood = -5.10    Transmembrane  321 - 337 ( 318 - 340)
INTEGRAL    Likelihood = -4.19    Transmembrane  203 - 219 ( 202 - 220)
INTEGRAL    Likelihood = -4.19    Transmembrane  244 - 260 ( 242 - 265)
INTEGRAL    Likelihood = -4.19    Transmembrane  284 - 300 ( 280 - 304)
INTEGRAL    Likelihood = -1.86    Transmembrane  177 - 193 ( 176 - 194)
```

```
----- Final Results -----
            bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF96429 GB:AE004383 conserved hypothetical protein [Vibrio cholerae]
 Identities = 144/442 (32%), Positives = 236/442 (52%), Gaps = 30/442 (6%)

Query: 18  NEFTFSNKSIIAYVWRGVVVGIIAGVIVSLFRLLIEVTADWVIEWYRYAHINSLLLLPIL 77
            N+F   +K+  + ++  ++VGI+AG++ + F   + + ++   +W + + I S L L +
Sbjct: 21  NQFLSKDKTPFSVLFLSLLVGILAGLVGTYFEQAVHLVSETRTDWLK-SEIGSFLPLWLA 79

Query: 78  SVSLLAVL-FVGFLV--KSDSDIKGSGIPHVEGELKGLMSPDWWSVLWKKFLGGIMAISM 134
            +  + A L F+G+ +  +   +  GSGIP +EG + G+    WW VL  KF GG+ A+
Sbjct: 80  AFLISAFLAFIGYFLVHRFAPEAAGSGIPEIEGAMDGMRPVRWWRVLPVKFFGGMGALGS 139

Query: 135 GFMLGREGPSIQLGAMSAKGLAKFLKSSRLEKR-VLIASGAAAGLSAAFNAPIAGLLFVV 193
            G +LGREGP++Q+G    +  ++   +    +  R  L+A+GAA GL+AAFNAP+AG++FV+
Sbjct: 140 GMVLGREGPTVQMGGAVGRMISDIFRVKNEDTRHSLLAAGAAGGLAAAFNAPLAGIMFVI 199

Query: 194 EEIYHHFS-RLIWITALVASLV-ANFISLNIFGLKPVLAMSEAMPFLGLNQYWLLLLLGL 251
            EE+   F   LI + A++ S V AN +   I G   V+ M +      L+   L LLLG
Sbjct: 200 EEMRPQFRYTLISVRAVIISAVAANIVFRVINGQDAVITMPQ-YDAPELSTLGLFLLLGA 258

Query: 252 FLGCLGYLYEIVIL-----------NFNKLYVILGSWLHLPDYFYGIIMVFLILPIGYYL 300
              G  G L+  +I            N  K Y++ GS +      +G++++        Y+
Sbjct: 259 LFGVFGVLFNYLITLAQDLFVKFHRNDRKRYLLTGSMI---GGCFGLLLL--------YV 307

Query: 301 PQLLGGGHGLILSLSNQQLPLMTIFFYFIIRFIVSMFSYGSGLPGGIFLPILTLGALAGL 360
            P+L GGG  LI +++N       +   F+ R   ++  +GSG PGGIF P+L LG L G
Sbjct: 308 PELTGGGISLIPTITNGGYGAGILLLLFVGRIFTTLLCFGSGAPGGIFAPMLALGTLFGY 367

Query: 361 LFGQIASQLGLLNQSFLSLFLILGMAGYFAAISKAPLTGMILVTEMVGDLKPLMAIAVVT 420
             FG IA            +F I GM   FAA  +AP+TG++LV EM  +   ++ + + +
Sbjct: 368 AFGLIAKMWFPELNIEPGMFAIAGMGALFAATVRAPITGILLVIEMTNNYHLILPLIITS 427

Query: 421 FVSYLVMDLLNGQPIYEAMLDK 442
              + +   LL GQPIY  +L +
Sbjct: 428 LGAVIFAQLLGGQPIYSQLLHR 449
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 343/510 (67%), Positives = 410/510 (80%)

Query: 1   MENHKNEFQFSLESILGFVWRGIVVGLIAGFVVSIFRLAIEKIFLVVMELYKSAHYQPII 60
            MENHKNEF FS +SI+ +VWRG+VVG+IAG +VS+FRL IE     V+E Y+ AH   ++
Sbjct: 13  MENHKNEFTFSNKSIIAYVWRGVVVGIIAGVIVSLFRLLIEVTADWVIEWYRYAHINSLL 72

Query: 61  LLSITVTSIIAAVIIGFFIKSDPDIKGSGIPHVEGELKGMLSPDWFSIVWKKFIAGILAI 120
            LL I   S++A + +GF +KSD DIKGSGIPHVEGELKG++SPDW+S++WKKF+ GI+AI
Sbjct: 73  LLPILSVSLLAVLFVGFLVKSDSDIKGSGIPHVEGELKGLMSPDWWSVLWKKFLGGIMAI 132

Query: 121 SSGLMLGREGPSIQLGAMTGKGIAQYLNASRMEKRVLIASGAAAGLSAAFNAPIAGLLFV 180
            S G MLGREGPSIQLGAM+ KG+A++L +SR+EKRVLIASGAAAGLSAAFNAPIAGLLFV
Sbjct: 133 SMGFMLGREGPSIQLGAMSAKGLAKFLKSSRLEKRVLIASGAAAGLSAAFNAPIAGLLFV 192

Query: 181 VEEIYHHFSRLVWITALVASLVANFVSLNIFGLTPVLALPSELPSLNLNFYWIFLLMGLF 240
```

```
           VEEIYHHFSRL+WITALVASLVANF+SLNIFGL PVLA+   +P L LN YW+ LL+GLF
Sbjct: 193 VEEIYHHFSRLIWITALVASLVANFISLNIFGLKPVLAMSEAMPFLGLNQYWLLLLLGLF 252

Query: 241 LGILGFIYEWVILRFHVIYDYLGKLFHLPSHLYGILAVIFILPIGYYFPQLLGGGNGLIV 300
           LG LG++YE VIL F+ +Y  LG   HLP + YGI+ V  ILPIGYY PQLLGGG+GLI+
Sbjct: 253 LGCLGYLYEIVILNFNKLYVILGSWLHLPDYFYGIIMVFLILPIGYYLPQLLGGGHGLIL 312

Query: 301 SLPRSNLSLMMLGLFFLIRFLWSMLSYSSGLPGGIFLPILALGSLAGAFFAVGMQYFGII 360
           SL    L LM +  +F+IRF+ SM SY SGLPGGIFLPIL LG+LAG  F       G++
Sbjct: 313 SLSNQQLPLMTIFFYFIIRFIVSMFSYGSGLPGGIFLPILTLGALAGLLFGQIASQLGLL 372

Query: 361 SHQQISLFVVLGMAGYFGAISKAPLTAMILVTEMVGDLKQLMAIGIVTMVSYIVMDLLKG 420
           +   +SLF++LGMAGYF AISKAPLT MILVTEMVGDLK LMAI +VT VSY+VMDLL G
Sbjct: 373 NQSFLSLFLILGMAGYFAAISKAPLTGMILVTEMVGDLKPLMAIAVVTFVSYLVMDLLNG 432

Query: 421 EPIYEAMLAKMTFNPKDKVMTPTLIELTVSDKISGKYVRDLELPENVLITTQIHHKTSAV 480
           +PIYEAML KM       ++ PTLIELTV DKI+GKYV++L+LPENVLITTQIHH+ S V
Sbjct: 433 QPIYEAMLDKMAMKHPTNLVEPTLIELTVGDKIAGKYVKELKLPENVLITTQIHHQKSQV 492

Query: 481 VSGNTILNAGDTIFLVVNESEIKEVREQLM 510
           VSGNT L +G TIFLVVNE++   VRE LM
Sbjct: 493 VSGNTRLLSGATIFLVVNEADTGFVREVLM 522
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1707

A DNA sequence (GBSx1811) was identified in *S.agalactiae* <SEQ ID 5305> which encodes the amino acid sequence <SEQ ID 5306>. This protein is predicted to be spermidine/putrescine-binding periplasmic protein precursor (potD-1). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.02   Transmembrane   20 -  36 (  14 -  40)

----- Final Results -----
              bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8881> which encodes amino acid sequence <SEQ ID 8882> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1    Crend: 2
SRCFLG: 0
McG: Length  of  UR:    22
     Peak Value of UR:    4.16
     Net Charge of CR: 2
McG: Discrim Score:     18.94
GvH: Signal Score (-7.5): -3.29
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program    count: 1 value:  -9.02 threshold:  0.0
   INTEGRAL    Likelihood = -9.02   Transmembrane    7 -  23 (   1 -  27)
   PERIPHERAL  Likelihood =  6.05     170
 modified ALOM score:   2.30
icml HYPID: 7  CFP: 0.461

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF94581 GB:AE004221 spermidine/putrescine ABC transporter,
          periplasmic spermidine/putrescine-binding protein [Vibrio cholerae]
 Identities = 126/327 (38%), Positives = 196/327 (59%), Gaps = 2/327 (0%)

Query: 42  SSSTPNSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMHTKIKQGGTTYDIA 101
           +++     +L  YNW +YI   +L+ FTKETGI+V Y T++SNE+M+ K+K  G  YD+
Sbjct: 18  TNAMAKDQELYFYNWSEYIPSEVLEDFTKETGIKVIYSTYESNESMYAKLKTQGAGYDLV 77

Query: 102 VPSDYMIDKMIKENLLVKLDHSKIANWDAIGARFKNLSFDPKNKYSIPYFWGTVGIVYN- 160
           VPS Y + KM KE +L ++DHSK++++  +   + N  FDP NK+SIPY WG  GI  N
Sbjct: 78  VPSTYFVSKMRKEGMLQEIDHSKLSHFKDLDPNYLNKPFDPGNKFSIPYIWGATGIGINT 137

Query: 161 DQLVKTPPKHWDDLWRPEFRNKIMLVDSAREVIGVGLNSLGYGLNTKNISELKAASKKLD 220
           D L K   K+W DLW  ++  ++ML+D AREV  + L+ LGY  NT N  E+KAA ++L
Sbjct: 138 DMLDKKSLKNWGDLWDAKWAGQLMLMDDAREVFHIALSKLGYSPNTTNPKEIKAAYRELK 197

Query: 221 ALTPNVKAIVADEMKGYMIQGDAAIGVTFSGEAREMLDGNKHLHYVVPSEGSNLWFDNIV 280
            L PNV    +D      + G+ ++G+ ++G A         +  + P +G+  W D+I
Sbjct: 198 KLMPNVLVFNSDFPANPYLAGEVSLGMLWNGSAYMARQEGAPIQIIWPEKGTIFWMDSIS 257

Query: 281 IPKTVKHRKEAYAFINFMMEPKNAAQNAEYIGYATPNLKAKALLPADIKNDKAFYPPDKT 340
           IP   K+ + A+  I+F++ P+NAA+ A  IGY TP   A  LLP +  ND + YPP
Sbjct: 258 IPAGAKNIEAAHKMIDFLLRPENAAKIALEIGYPTPVKTAHDLLPKEFANDPSIYPPQSV 317

Query: 341 IDHLEVYNNLGQKWLGIYNDLYLQFKM 367
           ID+ E  + +G+  + +Y++ + + K+
Sbjct: 318 IDNGEWQDEVGEASV-LYDEYFQKLKV 343
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5307> which encodes the amino acid sequence <SEQ ID 5308>. Analysis of this protein sequence reveals the following:

```
     Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.44    Transmembrane     8 -  24 (    1 -  27)

----- Final Results -----
                   bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC74207 GB:AE000212 spermidine/putrescine periplasmic transport
          protein [Escherichia coli]
 Identities = 134/342 (39%), Positives = 199/342 (58%), Gaps = 3/342 (0%)

Query: 17  ILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMYT 76
           +L + +  L   +     ++ L  YNW +Y+ P LL++FTKETGI+V Y T++SNE MY
Sbjct: 8   LLAAGALALGMSAAHADDNNTLYFYNWTEYVPPGLLEQFTKETGIKVIYSTYESNETMYA 67

Query: 77  KIKQ-GGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSL 135
           K+K      YD+ VPS Y +DKM KE ++ K+DKSKL    N+  + L K FDP NDYS+
Sbjct: 68  KLKTYKDGAYDLVVPSTYYVDKMRKEGMIQKIDKSKLTNFSNLDPDMLNKPFDPNNDYSI 127

Query: 136 PYFWGTVGIVYNDQLVD-KAPMHWEDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSK 194
           PY WG   I  N   VD K+   W DLW+PEYK S++L D ARE+  + L   GYS N+
Sbjct: 128 PYIWGATAIGVNGDAVDPKSVTSWADLWKPEYKGSLLLTDDAREVFQMALRKLGYSGNTT 187

Query: 195 NLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFSGEASEMLDSNEHLHYIV 254
           + ++++AA  +L++L PNV A  +D      ++G+  +G+ ++G A     +   +  +
Sbjct: 188 DPKEIEAAYNELKKLMPNVAAFNSDNPANPYMEGEVNLGMIWNGSAFVARQAGTPIDVVW 247

Query: 255 PSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGYATPNKKAKALLPDE 314
```

```
           P EG   W D+L +P   K+++ A   +NF+ RP+ A Q A  IGY TPN A+  LL  E
Sbjct: 248 PKEGGIFWMDSLAIPANAKNKEGALKLINFLLRPDVAKQVAETIGYPTPNLAARKLLSPE 307

Query: 315 IKNDPAFYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMYR 356
           + ND   YP  + IK  E  +++G+    IY + Y + K  R
Sbjct: 308 VANDKTLYPDAETIKNGEWQNDVGAA-SSIYEEYYQKLKAGR 348
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 270/357 (75%), Positives = 306/357 (85%)

Query: 14  MRRVYSFLGGIVLVILILFGLTTYLEKKSSSTPNSDKLVIYNWGDYIDPALLKKFTKETG 73
           MR++YSFL G++ VI+IL  L+  L+KKS S   SDKLVIYNWGDYIDPALLKKFTKETG
Sbjct: 1   MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETG 60

Query: 74  IEVQYETFDSNEAMHTKIKQGGTTYDIAVPSDYMIDKMIKENLLVKLDHSKIANWDAIGA 133
           IEVQYETFDSNEAM+TKIKQGGTTYDIAVPSDY IDKMIKENLL KLD SK+   D IG
Sbjct: 61  IEVQYETFDSNEAMYTKIKQGGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGK 120

Query: 134 RFKNLSFDPKNKYSIPYFWGTVGIVYNDQLVKTPPKHWDDLWRPEFRNKIMLVDSAREVI 193
            F   SFDP+N YS+PYFWGTVGIVYNDQLV   P HW+DLWRPE++N IML+D ARE++
Sbjct: 121 EFLGKSFDPQNDYSLPYFWGTVGIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREML 180

Query: 194 GVGLNSLGYGLNTKNISELKAASKKLDALTPNVKAIVADEMKGYMIQGDAAIGVTFSGEA 253
           GVGL + GY +N+KN+ +L+AA +KL  LTPNVKAIVADEMKGYMIQGDAAIG+TFSGEA
Sbjct: 181 GVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFSGEA 240

Query: 254 REMLDGNKHLHYVVPSEGSNLWFDNIVIPKTVKHRKEAYAFINFMMEPKNAAQNAEYIGY 313
            EMLD N+HLHY+VPSEGSNLWFDN+V+PKT+KH KEAYAF+NF+  P+NAAQNA YIGY
Sbjct: 241 SEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGY 300

Query: 314 ATPNLKAKALLPADIKNDKAFYPPDKTIDHLEVYNNLGQKWLGIYNDLYLQFKMYRK 370
           ATPN KAKALLP +IKND AFYP D  I  LEVY+NLG +WLGIYNDLYLQFKMYRK
Sbjct: 301 ATPNKKAKALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMYRK 357
```

SEQ ID 8882 (GBS135) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 6; MW 40kDa).

GBS135-His was purified as shown in Figure 201, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1708

A DNA sequence (GBSx1812) was identified in *S.agalactiae* <SEQ ID 5309> which encodes the amino acid sequence <SEQ ID 5310>. This protein is predicted to be spermidine/putrescine ABC transporter, permease protein (potC). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05    Transmembrane    17 -  33 (  10 -  37)
   INTEGRAL    Likelihood = -8.65    Transmembrane   236 - 252 ( 232 - 259)
   INTEGRAL    Likelihood = -7.75    Transmembrane   137 - 153 ( 132 - 158)
   INTEGRAL    Likelihood = -7.17    Transmembrane    63 -  79 (  60 -  92)
   INTEGRAL    Likelihood = -6.32    Transmembrane   108 - 124 ( 107 - 136)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8883> which encodes amino acid sequence <SEQ ID 8884> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1    Crend: 2
SRCFLG: 0
McG: Length  of  UR:   26
     Peak Value of UR:   3.65
     Net Charge of CR: 2
McG: Discrim Score:     16.58
GvH: Signal Score (-7.5): -6.17
     Possible site: 43
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 4 value: -12.05 threshold:  0.0
   INTEGRAL    Likelihood =-12.05   Transmembrane    9 -  25 (   2 -  29)
   INTEGRAL    Likelihood = -7.75   Transmembrane  129 - 145 ( 124 - 150)
   INTEGRAL    Likelihood = -7.17   Transmembrane   55 -  71 (  52 -  84)
   INTEGRAL    Likelihood = -6.32   Transmembrane  100 - 116 (  99 - 128)
   PERIPHERAL  Likelihood =  0.53     174
 modified ALOM score:   2.91
icm1 HYPID: 7  CFP: 0.582

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB91527 GB:AE001165 spermidine/putrescine ABC transporter,
          permease protein (potC) [Borrelia burgdorferi]
 Identities = 97/249 (38%), Positives = 159/249 (62%), Gaps = 3/249 (1%)

Query: 10  KKFANIYLALVFIILYIPIIYLIFYSFNKGGDMNSFTGFTFSHYGELFQDSRLMLILVQT 69
            + F NI+L L+    +Y+PII LI YSFN G      + GF+    Y E+F   S++    +  T
Sbjct: 3   RAFKNIFLFLILSFIYLPIIILIIYSFNSGDSGFIWQGFSLKWYKEIFASSQIKSAIFNT 62

Query: 70  FFLAFLSALLATIIGTFGAIWIYQVRRRH-QTSILSLNNILLVAPDVMIGASFLLVFTVI 128
              +A +S+L + +IG  GA  IY+    +  +T +LS+N I ++ PD++ G S +  ++ I
Sbjct: 63  ILIAIISSLTSVVIGIIGAYAIYKSENKKLKTILLSVNKITIINPDIVTGISLMTFYSAI 122

Query: 129 GLQLGFTSVLLSHVAFSIPIVVLMVLPRLKEMNDDMINASYDLGASTWQMLKEVMLPYLS 188
             +QLGF+++L+SH+ FS P VV+++LP+L  +  ++I+A+ DLGAS  Q+   ++ P ++
Sbjct: 123 KMQLGFSTMLISHIIFSTPYVVIIILPKLYSLPKNIIDAAKDLGASEIQIFFNIIYPEIA 182

Query: 189 SGIISGFFMAFTYSLDDFAVTFFVTGNGFSTLSVEIYSRARRGISLEINALSTIVF--LF 246
              I +G  +AFT S+DDF ++FF TG GF+ LS+ I S  +RGI   INA+S I+F  +
Sbjct: 183 GSIATGALIAFTLSIDDFLISFFTTGQGFNNLSILINSLTKRGIKPVINAISAILFFTIL 242

Query: 247 SILLVIGYY 255
            S+L +I  +
Sbjct: 243 SLLFIINKF 251
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5311> which encodes the amino acid sequence <SEQ ID 5312>. Analysis of this protein sequence reveals the following:

```
     Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.17   Transmembrane    9 -  25 (   4 -  29)
   INTEGRAL    Likelihood = -8.12   Transmembrane  228 - 244 ( 224 - 250)
   INTEGRAL    Likelihood = -7.91   Transmembrane  129 - 145 ( 124 - 150)
   INTEGRAL    Likelihood = -7.06   Transmembrane   62 -  78 (  54 -  87)
   INTEGRAL    Likelihood = -3.93   Transmembrane  100 - 116 (  99 - 118)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB91527 GB:AE001165 spermidine/putrescine ABC transporter,
          permease protein (potC) [Borrelia burgdorferi]
 Identities = 91/249 (36%), Positives = 154/249 (61%), Gaps = 3/249 (1%)

Query: 2   KKFANLYLASVFVLLYIPIFYLIFYSFNKGGDMNGFTGFTLEHYQTMFEDSRLMTILLQT 61
           + F N++L  +   +Y+PI  LI YSFN G     + GF+L+ Y+ +F  S++ + +  T
Sbjct: 3   RAFKNIFLFLILSFIYLPIIILIIYSFNSGDSGFIWQGFSLKWYKEIFASSQIKSAIFNT 62

Query: 62  FVLAFSSALLATIIGIFGAIFIHHVRGK-YQNAMLSANNVLMVSPDVMIGASFLILFTSL 120
            ++A  S+L + +IGI GA  I+    K  +  +LS N + +++PD++ G S +  ++++
Sbjct: 63  ILIAIISSLTSVVIGIIGAYAIYKSENKKLKTILLSVNKITIINPDIVTGISLMTFYSAI 122

Query: 121 KFQLGMSSVLLSHIAFSIPIVVLMVLPRLKEMNQDMVNAAYDLGANYFQMLKEVMLPYFT 180
           K QLG S++L+SHI FS P VV+++LP+L  + +++++AA DLGA+  Q+   ++ P
Sbjct: 123 KMQLGFSTMLISHIIFSTPYVVIIILPKLYSLPKNIIDAAKDLGASEIQIFFNIIYPEIA 182

Query: 181 PGIIAGYFMAFTYSLDDFAVTFFLTGNSVTTLSVEIYSRARQGISLDINALSTIVFF--F 238
             I  G  +AFT S+DDF ++FF TG     LS+ I S  ++GI   INA+S I+FF
Sbjct: 183 GSIATGALIAFTLSIDDFLISFFTTGQGFNNLSILINSLTKRGIKPVINAISAILFFTIL 242

Query: 239 SILLVIGYY 247
           S+L +I  +
Sbjct: 243 SLLFIINKF 251
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 196/258 (75%), Positives = 231/258 (88%)

Query: 9   MKKFANIYLALVFIILYIPIIYLIFYSFNKGGDMNSFTGFTFSHYGELFQDSRLMLILVQ 68
           MKKFAN+YLA VF++LYIPI YLIFYSFNKGGDMN FTGFT  HY  +F+DSRLM IL+Q
Sbjct: 1   MKKFANLYLASVFVLLYIPIFYLIFYSFNKGGDMNGFTGFTLEHYQTMFEDSRLMTILLQ 60

Query: 69  TFFLAFLSALLATIIGTFGAIWIYQVRRRHQTSILSLNNILLVAPDVMIGASFLLVFTVI 128
           TF LAF SALLATIIG FGAI+I+ VR ++Q ++LS NN+L+V+PDVMIGASFL++FT +
Sbjct: 61  TFVLAFSSALLATIIGIFGAIFIHHVRGKYQNAMLSANNVLMVSPDVMIGASFLILFTSL 120

Query: 129 GLQLGFTSVLLSHVAFSIPIVVLMVLPRLKEMNDDMINASYDLGASTWQMLKEVMLPYLS 188
             QLG +SVLLSH+AFSIPIVVLMVLPRLKEMN DM+NA+YDLGA+ +QMLKEVMLPY +
Sbjct: 121 KFQLGMSSVLLSHIAFSIPIVVLMVLPRLKEMNQDMVNAAYDLGANYFQMLKEVMLPYFT 180

Query: 189 SGIISGFFMAFTYSLDDFAVTFFVTGNGFSTLSVEIYSRARRGISLEINALSTIVFLFSI 248
            GII+G+FMAFTYSLDDFAVTFF+TGN  +TLSVEIYSRAR+GISL+INALSTIVF FSI
Sbjct: 181 PGIIAGYFMAFTYSLDDFAVTFFLTGNSVTTLSVEIYSRARQGISLDINALSTIVFFFSI 240

Query: 249 LLVIGYYYISKEKGEKNA 266
           LLVIGYYY+S++K EK+A
Sbjct: 241 LLVIGYYYMSQDKEEKHA 258
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1709

A DNA sequence (GBSx1813) was identified in *S.agalactiae* <SEQ ID 5313> which encodes the amino acid sequence <SEQ ID 5314>. This protein is predicted to be spermidine/putrescine ABC transporter, permease protein (potB). Analysis of this protein sequence reveals the following:

```
Possible site: 35
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.55   Transmembrane  250 - 266 ( 244 - 269)
   INTEGRAL    Likelihood = -3.93   Transmembrane  148 - 164 ( 146 - 166)
   INTEGRAL    Likelihood = -3.35   Transmembrane   65 -  81 (  64 -  85)
   INTEGRAL    Likelihood = -1.97   Transmembrane   96 - 112 (  96 - 115)

----- Final Results -----
               bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9853> which encodes amino acid sequence <SEQ ID 9854> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22990 GB:U32813 spermidine/putrescine ABC transporter,
           permease protein (potB) [Haemophilus influenzae Rd]
 Identities = 90/255 (35%), Positives = 153/255 (59%), Gaps = 11/255 (4%)

Query: 21  AWLFLFVLAPVALIAWNSFFDINGH------FTLANYQTFFSSGTYLKMSFNSVLYAGIV 74
            +WL  FVL P  L+   SF   +G        T+ NY   F+   Y ++ +NS+  +GI
Sbjct: 18  SWLIFFVLIPNLLVLAVSFLTRDGSNFYAFPITIENYTNLFNP-LYAQVVWNSLSMSGIA 76

Query: 75  SFITLLISYPAAYLLTKL--KHKQLWLMLVILPTWINLLLKAYAFMGIFGQQGGINAFLT 132
            + I LLI YP A++++K+  K++ L L LV+LP W N L++ Y      G +G +N  L
Sbjct: 77  TIICLLIGYPFAFMMSKIHPKYRPLLLFLVVLPFWTNSLIRIYGMKVFLGVKGILNTMLI 136

Query: 133 FIGI--GPKQILFTDFSFLFVAAYIELPFMLLPIFNALDDIDQNLIYASDDLGANAWQTF 190
             +GI   P +IL T+ + +    Y+ LPFM+LP+++A++ +D  L+ A+ DLGAN +Q F
Sbjct: 137 DMGILSAPIRILNTEIAVIIGLVYLLLPFMILPLYSAIEKLDNRLLEAARDLGANTFQRF 196

Query: 191 QKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLITQNKGMGST 250
             +VI PL++ G+ AG   V +P++ +F +  L+GG +V+ +G  I+  FLI++N   GS
Sbjct: 197 FRVILPLTMPGIIAGCLLVLLPAMGMFYVADLLGGAKVLLVGNVIKSEFLISRNWPFGSA 256

Query: 251 IGVILILVMVAIMWL 265
            + + L ++M  ++++
Sbjct: 257 VSIGLTVLMALLIFV 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5315> which encodes the amino acid sequence <SEQ ID 5316>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.38   Transmembrane   19 -  35 (  11 -  40)
   INTEGRAL    Likelihood = -6.79   Transmembrane  250 - 266 ( 245 - 268)
   INTEGRAL    Likelihood = -4.83   Transmembrane   65 -  81 (  63 -  85)
   INTEGRAL    Likelihood = -1.97   Transmembrane   96 - 112 (  96 - 115)
   INTEGRAL    Likelihood = -1.91   Transmembrane  148 - 164 ( 148 - 165)

----- Final Results -----
               bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC22990 GB:U32813 spermidine/putrescine ABC transporter,
           permease protein (potB) [Haemophilus influenzae Rd]
 Identities = 91/262 (34%), Positives = 158/262 (59%), Gaps = 11/262 (4%)

Query: 20  FLWILFFVVAPVTLLFYKSFFDIEGR------VTLANYETFFSSWTYLRMSVNSILYAGI 73
            F W++FFV+ P  L+   SF   +G       +T+ NY   F+   Y ++  NS+  +GI
Sbjct: 17  FSWLIFFVLIPNLLVLAVSFLTRDGSNFYAFPITIENYTNLFNP-LYAQVVWNSLSMSGI 75
```

```
Query: 74  ITLVTLLISYPTALFLTRL--KHKQLWLMLIILPTWVNLLLKAYAFMGIFGQQGGINSFL 131
            T++ LLI YP A  ++++  K++ L L L++LP W N L++ Y      G +G +N+ L
Sbjct: 76  ATIICLLIGYPFAFMMSKIHPKYRPLLLFLVVLPFWTNSLIRIYGMKVFLGVKGILNTML 135

Query: 132 TFMGI--GPQQILFTDFSFIFVASYIELPFMMLPIFNALDDIDHNVINASRDLGASEFQA 189
              MGI   P +IL T+ + I    Y+ LPFM+LP+++A++ +D+ ++ A+RDLGA+ FQ
Sbjct: 136 IDMGILSAPIRILNTEIAVIIGLVYLLLPFMILPLYSAIEKLDNRLLEAARDLGANTFQR 195

Query: 190 FSKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLTTQNWGMGS 249
           F +VI PL++ G+ AG   V +P++ +F +  L+GG +V+ +G  I+  FL ++NW  GS
Sbjct: 196 FFRVILPLTMPGIIAGCLLVLLPAMGMFYVADLLGGAKVLLVGNVIKSEFLISRNWPFGS 255

Query: 250 TIGVVLILTMVAIMWLTKEKSK 271
            + + L + M  ++++    +K
Sbjct: 256 AVSIGLTVLMALLIFVYYRANK 277
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 215/266 (80%), Positives = 239/266 (89%)

Query: 4   RRREMKKTSSLFSIPYMAWLFLFVLAPVALIAWNSFFDINGHFTLANYQTFFSSGTYLKM 63
           RR  MKKTSSLFSIPY  W+  FV+APV L+ + SFFDI G  TLANY+TFFSS TYL+M
Sbjct: 4   RRSVMKKTSSLFSIPYFLWILFFVVAPVTLLFYKSFFDIEGRVTLANYETFFSSWTYLRM 63

Query: 64  SFNSVLYAGIVSFITLLISYPAAYLLTKLKHKQLWLMLVILPTWINLLLKAYAFMGIFGQ 123
           S NS+LYAGI++ +TLLISYP A  LT+LKHKQLWLML+ILPTW+NLLLKAYAFMGIFGQ
Sbjct: 64  SVNSILYAGIITLVTLLISYPTALFLTRLKHKQLWLMLIILPTWVNLLLKAYAFMGIFGQ 123

Query: 124 QGGINAFLTFIGIGPKQILFTDFSFLFVAAYIELPFMLLPIFNALDDIDQNLIYASDDLG 183
           QGGIN+FLTF+GIGP+QILFTDFSF+FVA+YIELPFM+LPIFNALDDID N+I AS DLG
Sbjct: 124 QGGINSFLTFMGIGPQQILFTDFSFIFVASYIELPFMMLPIFNALDDIDHNVINASRDLG 183

Query: 184 ANAWQTFQKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLITQ 243
           A+ +Q F KVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFL TQ
Sbjct: 184 ASEFQAFSKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLTTQ 243

Query: 244 NKGMGSTIGVILILVMVAIMWLTKER 269
           N GMGSTIGV+LIL MVAIMWLTKE+
Sbjct: 244 NWGMGSTIGVVLILTMVAIMWLTKEK 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1710

A DNA sequence (GBSx1814) was identified in *S.agalactiae* <SEQ ID 5317> which encodes the amino acid sequence <SEQ ID 5318>. This protein is predicted to be spermidine/putrescine ABC transporter, ATP-binding protein (potA). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3031(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB91525 GB:AE001165 spermidine/putrescine ABC transporter,
           ATP-binding protein (potA) [Borrelia burgdorferi]
 Identities = 166/345 (48%), Positives = 240/345 (69%), Gaps = 1/345 (0%)

Query: 1   MTNPIIAFKNVSKVFEDSNTVVLKDINFELEEGKFYTLLGASGSGKSTILNIIAGLLEAS 60
           M N I+  KN+S  ++++    L +IN ++++ +F TLLG SG GK+T++ I+ G L
```

```
Sbjct: 1   MDNCILEIKNLSHYYDNNGNKTLDNINLKIKKNEFITLLGPSGCGKTTLIKILGGFLSQK 60

Query: 61  TGDIYLDGKRINDVPTNKRDVHTVFQNYALFPHMTVFENVAFPLKLKKMDKKEIQKRVQE 120
            G+IY   K I+    NKR+++TVFQNYALFPHM VF+N++F L++KK  K  I+++V+
Sbjct: 61  NGEIYFFSKEISKTSPNKREINTVFQNYALFPHMNVFDNISFGLRMKKTPKDIIKEKVKT 120

Query: 121 TLKMVRLEGFEKRAIQKLSGGQRQRVAIARAIINQPKVVLLDEPLSALDLKLRTEMQYEL 180
           +L ++ +  +  R I +LSGGQ+QRVAIARA++ +PK++LLDEPLSALDLK+R EMQ EL
Sbjct: 121 SLSLIGMPKYAYRNINELSGGQKQRVAIARAMVMEPKLLLLDEPLSALDLKMRQEMQKEL 180

Query: 181 RELQQRLGITFVFVTHDQEEALAMSDWIFVMNEGEIVQSGTPVDIYDEPINHFVATFIGE 240
           +++Q++LGITF++VTHDQEEAL MSD I VMNEG I+Q GTP +IY+EP   FVA FIGE
Sbjct: 181 KKIQRQLGITFIYVTHDQEEALTMSDRIVVMNEGIILQIGTPEEIYNEPKTKFVADFIGE 240

Query: 241 SNILSGKMIEDYLVEFNGKRFEAVDGGMRPNESVQVVIRPEDLQITLPDEGKLQVKVDTQ 300
           SNI  G   ++ +V   G  FE +D G    E+V +VIRPED+++    +G L   + +
Sbjct: 241 SNIFDGTYKKELVVSLLGHEFECLDKGFEAEEAVDLVIRPEDVKLLPKGKGHLSGTITSA 300

Query: 301 LFRGVHYEIIAYDDLGNEWMIHSTRKAIEGEVIGLDFTPEDIHIM 345
           +F+GVHYE+       N W++ STR    GE + +   P+DIH+M
Sbjct: 301 IFQGVHYEMTLEIQKTN-WIVQSTRLTKVGEEVDIFLEPDDIHVM 344
```

There is also homology to SEQ ID 1292

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1711

A DNA sequence (GBSx1815) was identified in *S.agalactiae* <SEQ ID 5319> which encodes the amino acid sequence <SEQ ID 5320>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4990(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06283 GB:AP001515 UDP-N-acetylenolpyruvoylglucosamine
           reductase [Bacillus halodurans]
 Identities = 119/286 (41%), Positives = 166/286 (57%), Gaps = 1/286 (0%)

Query: 13  DIRFDEPLKKYTYTKVGGPADYLAFPRNRLELSRIVKFANSQNIPWMVLGNASNIIVRDG 72
           ++R +E L  +T  K+GGPAD    P +   L   +K        W V+G  SNI+V D
Sbjct: 15  EVRVNESLAHHTTWKIGGPADVFVIPNDIEGLKNTMKLIQETGCKWRVIGRGSNILVSDK 74

Query: 73  GIRGFVIMFDK-LSTVTVNGYVIEAEAGANLIETTRIARYHSLTGFEFACGIPGSVGGAV 131
           G+RG  I  DK L  + VNG  I   AG  +++   +     L G EFA GIPGSVGGAV
Sbjct: 75  GLRGVTIKLDKGLDHLEVNGESITVGAGFPVVKLATVISRQGLAGLEFAAGIPGSVGGAV 134

Query: 132 FMNAGAYGGEIAHILLSAQVLTPQGELKTIEARNMQFGYRHSVIQESGDIVISAKFALKP 191
           FMNAGA+G +I+ IL  A VL P G L+ +    M F YR S++Q++  I + A F+L
Sbjct: 135 FMNAGAHGSDISQILTKAHVLFPDGTLRWLTNEEMAFSYRTSLLQKNDGICVEAIFSLTR 194

Query: 192 GDHLMITQEMDRLTYLRELKQPLEYPSCGSVFKRPPGHFAGQLISEAHLKGQRIGGVEVS 251
           GD   I +++ +    R   QP  +P+CGSVF+ P   +AGQLI +A LKG +IGG ++S
Sbjct: 195 GDKEDIKKKLQKNKDYRRDTQPWNHPTCGSVFRNPLPEYAGQLIEKAGLKGYQIGGAQIS 254

Query: 252 QKHAGFMVNIAEGSAQDYENLIEHVINTVESTSGVHLEPEVRIIGE 297
             HA F+VN  +  A D   LI HV +T++    +++E EV +IGE
Sbjct: 255 TMHANFIVNTGDAKAADVLALIHHVKDTIQKQYQMNMETEVELIGE 300
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5321> which encodes the amino acid sequence <SEQ ID 5322>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4557(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 229/292 (78%), Positives = 267/292 (91%)

Query: 8   ELEGLDIRFDEPLKKYTYTKVGGPADYLAFPRNRLELSRIVKFANSQNIPWMVLGNASNI 67
           EL G+DIR +EPLK YTYTKVGGPAD+LAFPRN  ELSRIV +AN +N+PW+VLGNASN+
Sbjct: 4   ELHGIDIRENEPLKHYTYTKVGGPADFLAFPRNHYELSRIVAYANKENMPWLVLGNASNL 63

Query: 68  IVRDGGIRGFVIMFDKLSTVTVNGYVIEAEAGANLIETTRIARYHSLTGFEFACGIPGSV 127
           IVRDGGIRGFVIMFDKL+ V +NGY +EAEAGANLIETT+IA++HSLTGFEFACGIPGS+
Sbjct: 64  IVRDGGIRGFVIMFDKLNAVHLNGYTLEAEAGANLIETTKIAKFHSLTGFEFACGIPGSI 123

Query: 128 GGAVFMNAGAYGGEIAHILLSAQVLTPQGELKTIEARNMQFGYRHSVIQESGDIVISAKF 187
           GGAVFMNAGAYGGEI+HI LSA+VLTP GE+KTI AR+M FGYRHS IQE+GDIVISAKF
Sbjct: 124 GGAVFMNAGAYGGEISHIFLSAKVLTPSGEIKTISARDMAFGYRHSAIQETGDIVISAKF 183

Query: 188 ALKPGDHLMITQEMDRLTYLRELKQPLEYPSCGSVFKRPPGHFAGQLISEAHLKGQRIGG 247
           ALKPG++  I+QEM+RL +LR+LKQPLE+PSCGSVFKRPPGHFAGQLI EA+LKG RIGG
Sbjct: 184 ALKPGNYDTISQEMNRLNHLRQLKQPLEFPSCGSVFKRPPGHFAGQLIMEANLKGHRIGG 243

Query: 248 VEVSQKHAGFMVNIAEGSAQDYENLIEHVINTVESTSGVHLEPEVRIIGESL 299
           VEVS+KH GFM+N+A+G+A+DYE+LI +VI TVE+ SGV LEPEVRIIGE+L
Sbjct: 244 VEVSEKHTGFMINVADGTAKDYEDLIAYVIETVENHSGVRLEPEVRIIGENL 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1712

A DNA sequence (GBSx1816) was identified in *S.agalactiae* <SEQ ID 5323> which encodes the amino acid sequence <SEQ ID 5324>. This protein is predicted to be 2-amino-4-hydroxy-6-hydroxymethyldihydropterin pyrophosphokinase/dihyd. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1122(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03814 GB:AP001507
           2-amino-4-hydroxy-6-hydroxymethyldihydropteridin e
           pyrophosphokinase [Bacillus halodurans]
 Identities = 64/146 (43%), Positives = 94/146 (63%)

Query: 5   YLSLGSNIGDRETFLKQALFSIDHLQKTKVAQISAIYETAAWGNTNQEDFFNICCQVETD 64
           Y++LGSNIGDR  FL++A+  +    K  V   S+IYET   G T+Q  F N+  +V T
Sbjct: 6   YIALGSNIGDRSRFLEEAIQQLAEHDKVTVTCCSSIYETDPVGYTDQSPFLNMVVEVSTS 65
```

```
Query: 65  LAPFELLDYCQEIEKCLKRVRHEHWGPRTIDIDILLFGNQVINQEDLVVPHPYMTKRAFV 124
           L   +LL+  Q+IE+   R RH  WGPRT+D+DILL+  +    E+L++PHP M +RAFV
Sbjct: 66  LPVEQLLEVTQKIERYCGRERHIRWGPRTLDLDILLYDQENREMENLIIPHPRMWERAFV 125

Query: 125 LVPLLEIAPQLSLPNGSKLEDYLEKL 150
           L+PL+E+ P +  P+G  +E  + +L
Sbjct: 126 LIPLMELNPSIVAPSGKTIEQVVREL 151
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5325> which encodes the amino acid sequence <SEQ ID 5326>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0479(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 85/156 (54%), Positives = 111/156 (70%), Gaps = 1/156 (0%)

Query: 1   MTTVYLSLGSNIGDRETFLKQALFSIDHLQKTKVAQISAIYETAAWGNTNQEDFFNICCQ 60
           MT VYLSLG+N+GDR  +L++AL ++  L +T++   S+IYET AWG T Q DF N+ CQ
Sbjct: 1   MTIVYLSLGTNMGDRAAYLQKALEALADLPQTRLLAQSSIYETTAWGKTGQADFLNMACQ 60

Query: 61  VETDLAPFELLDYCQEIEKCLKRVRHEHWGPRTIDIDILLFGNQVINQEDLVVPHPYMTK 120
           ++T L   + L   Q IE+ L RVRHE WG RTIDIDILLFG +V + ++L VPHPYMT+
Sbjct: 61  LDTQLTAADFLKETQAIEQSLGRVRHEKWGSRTIDIDILLFGEEVYDTKELKVPHPYMTE 120

Query: 121 RAFVLVPLLEIAPQLSLPNGSK-LEDYLEKLNLGEV 155
           RAFVL+PLLE+ P L LP   K L DYL  L+  ++
Sbjct: 121 RAFVLIPLLELQPDLKLPPNHKFLRDYLAALDQSDI 156
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1713

A DNA sequence (GBSx1817) was identified in *S.agalactiae* <SEQ ID 5327> which encodes the amino acid sequence <SEQ ID 5328>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2826(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5329> which encodes the amino acid sequence <SEQ ID 5330>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3547(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 75/119 (63%), Positives = 92/119 (77%)

Query: 1   MDKIYLNKCRFYGYHGAFSEEQTLGQVFQVDAVLSLDLAKASQTDDLIDTVHYGEVFDCI 60
           MDKI L  CRFYGYHGAF EEQTLGQ+F VD  LS+DL  AS +D L DTVHYG VFD +
Sbjct: 1   MDKIVLEGCRFYGYHGAFKEEQTLGQIFLVDLELSVDLQAASLSDQLTDTVHYGMVFDSV 60

Query: 61  KNHVENEQYQLIEKLAGVIVEDIFLQFHPVQAITLKITKDNPPINGHYESVGIELERRR 119
           +  VE E++ LIE+LAG I E +F +F P++AI + I K+NPPI GHY++VGIELER+R
Sbjct: 61  RQLVEGEKFILIERLAGAICEQLFNEFPPIEAIKVAIKKENPPIAGHYKAVGIELERQR 119
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1714

A DNA sequence (GBSx1818) was identified in *S.agalactiae* <SEQ ID 5331> which encodes the amino acid sequence <SEQ ID 5332>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5333> which encodes the amino acid sequence <SEQ ID 5334>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 181/267 (67%), Positives = 224/267 (83%), Gaps = 1/267 (0%)

Query: 1   MKIGQYDITGKACIMGILNVTPDSFSDGGSYTTIDSALNQVGEMLEQGVAIVDIGGESTR 60
           MKIG++ I G A IMGILNVTPDSFSDGGSYTT+  AL+ V +M+  G  I+D+GGESTR
Sbjct: 1   MKIGKFVIEGNAAIMGILNVTPDSFSDGGSYTTVQKALDHVEQMIADGAKIIDVGGESTR 60

Query: 61  PGAVFVTAEEEIKRVVPMIKAIREVYPDLLLSIDTYKTEVAQAALDAGVHILNDVWSGLY 120
           PG  FV+A +EI RVVP+IKAI+E Y D+L+SIDTYKTE A+AAL+AG  ILNDVW+GLY
Sbjct: 61  PGCQFVSATDEIDRVVPVIKAIKENY-DILISIDTYKTETARAALEAGADILNDVWAGLY 119

Query: 121 DGKMLSLAAERNVPIILMHNQEEAVYQDIKKEVCEFLLERAERALEAGVSKDNIWIDPGF 180
           DG+M +LAAE + PIILMHNQ+E VYQ++ ++VC+FL  RA+ AL+AGV K+NIW+DPGF
Sbjct: 120 DGQMFALAAEYDAPIILMHNQDEEVYQEVTQDVCDFLGNRAQAALDAGVPKNNIWVDPGF 179

Query: 181 GFAKTEEQNLELLKGLEQVCDLGYPVLFGISRKRTVNYLLGGNREVTERDMGTAALSAWA 240
           GFAK+ +QN ELLKGL++VC LGYPVLFGISRKR V+ LLGGN +  ERD  TAALSA+A
Sbjct: 180 GFAKSVQQNTELLKGLDRVCQLGYPVLFGISRKRVVDALLGGNTKAKERDGATAALSAYA 239

Query: 241 IAKGCQIVRVHNVEVNKDIVTVISQLV 267
           + KGCQIVRVH+V+ N+DIV V+SQL+
Sbjct: 240 LGKGCQIVRVHDVKANQDIVAVLSQLM 266
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1715

A DNA sequence (GBSx1819) was identified in *S.agalactiae* <SEQ ID 5335> which encodes the amino acid sequence <SEQ ID 5336>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2429(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5337> which encodes the amino acid sequence <SEQ ID 5338>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1590(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 151/184 (82%), Positives = 166/184 (90%)

Query: 3   NQEKMEKAIYQFLEALGENPNREGLKDTPKRVAKMYIEMFSGLNQDPKEQFTAVFSENHE 62
            N+EK E AIYQFLEA+GENPNREGL DTPKRVAKMY EMF GL +DPKE+FTAVF E HE
Sbjct: 16  NKEKAEAAIYQFLEAIGENPNREGLLDTPKRVAKMYAEMFLGLGKDPKEEFTAVFKEQHE 75

Query: 63  EVVIVKDIPFYSMCEHHLVPFYGKAHIAYLPNDGRVTGLSKLARAVEVASKRPQLQERLT 122
            +VVIVKDI FYS+CEHHLVPFYGKAHIAYLP+DGRVTGLSKLARAVEVASKRPQLQERLT
Sbjct: 76  DVVIVKDISFYSICEHHLVPFYGKAHIAYLPSDGRVTGLSKLARAVEVASKRPQLQERLT 135

Query: 123 AQVAQALEDALAPKGIFVMIEAEHMCMTMRGIKKPGSKTITTVARGLYKDDRYERQEILS 182
            +Q+A AL +AL PKG  VM+EAEHMCMTMRGIKKPGSKTITT ARGLYK+ R ERQE++S
Sbjct: 136 SQIADALVEALNPKGTLVMVEAEHMCMTMRGIKKPGSKTITTTARGLYKESRAERQEVIS 195

Query: 183 LIQK 186
            L+ K
Sbjct: 196 LMTK 199
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1716

A DNA sequence (GBSx1820) was identified in *S.agalactiae* <SEQ ID 5339> which encodes the amino acid sequence <SEQ ID 5340>. This protein is predicted to be folylpolyglutamate synthase (folC). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2836(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9855> which encodes amino acid sequence <SEQ ID 9856> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14768 GB:Z99118 folyl-polyglutamate synthetase [Bacillus subtilis]
 Identities = 154/426 (36%), Positives = 245/426 (57%), Gaps = 17/426 (3%)

Query: 3   YQEALEWIHSKLAFGIKPGLERMRWMLEQLGNPQNNLSAIHVVGTNGKGSTTSYLQHIFT 62
           YQ+A  WIH +L FG+KPGL RM+ ++ +LG+P+  + A HV GTNGKGST ++++ +
Sbjct: 5   YQDARSWIHGRLKFGVKPGLGRMKQLMARLGHPEKKIRAFHVAGTNGKGSTVAFIRSMLQ 64

Query: 63  NSGYQVGTFTSPYIVDFRERISIDGQMIPESDFIKLVETVRPVVERLHLETNLEPATEFE 122
            +GY VGTFTSPYI+ F ERIS++G  I + ++  LV  ++P VE L  +T     TEFE
Sbjct: 65  EAGYTVGTFTSPYIITFNERISVNGIPISDEEWTALVNQMKPHVEALD-QTEYGQPTEFE 123

Query: 123 VITVLMFYYFGNSCPVDIVIIEAGMGGYYDSTNMFKALAVTCPSIGLDHQEVLGRTYVDI 182
           ++T   F YF     VD VI E G+GG +DSTN+ + L      SIG DH  +LG T  +I
Sbjct: 124 IMTACAFLYFAEFHKVDFVIFETGLGGRFDSTNVVEPLLTVITSIGHDHMNILGNTIEEI 183

Query: 183 AEQKVGVLKKGVPFVYANDRQDVEEVFQIKAKETHSQTYRLHNDFYIKEEE-----NYFN 237
           A +K G++K+G+P V A  + +  +V + +A+   +    LH+   I  EE       F+
Sbjct: 184 AGEKAGIIKEGIPIVTAVTQPEALQVIRHEAERHAAPFQSLHDACVIFNEEALPAGEQFS 243

Query: 238 YIGPQANIDHIQLQMPGHHQVSNASIAI-TTSLLLRDKYPKLTLQTIKDGLEMTKWVGRT 296
           +   +   + I+  + G HQ  NA+++I     L ++    ++ + ++ GL    W GR
Sbjct: 244 FKTEEKCYEDIRTSLIGTHQRQNAALSILAAEWLNKENIAHISDEALRSGLVKAAWPGRL 303

Query: 297 ELI--FPNVMIDGAHNNESVDALVQVIK-KYQQKNVHILFAAINTKPIESMLESLSSIA- 352
           EL+   P V +DGAHN E V+ L + +K ++    + ++F+A+  KP ++M++ L +IA
Sbjct: 304 ELVQEHPPVYLDGAHNEEGVEKLAETMKQRFANSRISVVFSALKDKPYQNMIKRLETIAH 363

Query: 353 PVSVTSFDYPK-SINLDKYPKAYTRVSDWKKWLHDI-----NLTSDKDFYVITGSLYFIS 406
            +   SFD+P+ S+  D Y  +      W +   D+     +     +  +ITGSLYFIS
Sbjct: 364 AIHFASFDFPRASLAKDLYDASEISNKSWSEDPDDVIKFIESKKGSNEIVLITGSLYFIS 423

Query: 407 QVRQEL 412
            +R+ L
Sbjct: 424 DIRKRL 429
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5341> which encodes the amino acid sequence <SEQ ID 5342>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.28   Transmembrane   12 -  28 (  12 -  28)

----- Final Results -----
              bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 230/411 (55%), Positives = 295/411 (70%), Gaps = 1/411 (0%)

Query: 1   MTYQEALEWIHSKLAFGIKPGLERMRWMLEQLGNPQNNLSAIHVVGTNGKGSTTSYLQHI 60
           MTY+E LEWIH  L FGIKPGL+RM W+L QLGNPQ N+  +H+VGTNGKGST ++LQHI
Sbjct: 34  MTYEETLEWIHDHLVFGIKPGLKRMLWVLGQLGNPQKNVKGVHIVGTNGKGSTVNHLQHI 93

Query: 61  FTNSGYQVGTFTSPYIVDFRERISIDGQMIPESDFIKLVETVRPVVERLHLETNLEPATE 120
           FT +GY+VGTFTSPYI+DF+ERISI+G+MI E D +     +RP+ ERL  ET+    TE
Sbjct: 94  FTTAGYEVGTFTSPYIMDFKERISINGRMISEKDLVIAANRIRPLTERLVQETDFGEVTE 153

Query: 121 FEVITVLMFYYFGNSCPVDIVIIEAGMGGYYDSTNMFKALAVTCPSIGLDHQEVLGRTYV 180
           FEVIT++MF YFG+  PVDI IIEAG+GG YDSTN+F+A+ V CPSIGLDHQ +LG TY
Sbjct: 154 FEVITLIMFLYFGDMHPVDIAIIEAGLGGLYDSTNVFQAMVVVCPSIGLDHQAILGETYA 213

Query: 181 DIAEQKVGVLKKGVPFVYANDRQDVEEVFQIKAKETHSQTYRLHNDFYIKEEENYFNYIG 240
           +IA QK GVL+ G   V+A +     EVF  KA++  +  +     F + E  + + +
```

```
Sbjct: 214 NIAAQKAGVLEGGETLVFAVENPSAREVFLTKAEQVGASIWEWQEQFQMAENASGYRFTS 273

Query: 241 PQANIDHIQLQMPGHHQVSNASIAITTSLLLRDKYPKLTLQTIKDGLEMTKWVGRTELIF 300
           P   I  I + MPGHHQVSNA++AI T L L+D+YP+LT   I++GL  + W+GRTEL+
Sbjct: 274 PLGVISDIHIAMPGHHQVSNAALAIMTCLTLQDRYPRLTPDHIREGLANSLWLGRTELLA 333

Query: 301 PNVMIDGAHNNESVDALVQVIK-KYQQKNVHILFAAINTKPIESMLESLSSIAPVSVTSF 359
           PN+MIDGAHNNESV ALV V+K  Y  K +HILF AI+TKPI  ML +L  I  + VTSF
Sbjct: 334 PNLMIDGAHNNESVAALVAVLKNNYNDKKLHILFGAIDTKPIADMLVALEQIGDLQVTSF 393

Query: 360 DYPKSINLDKYPKAYTRVSDWKKWLHDINLTSDKDFYVITGSLYFISQVRQ 410
            YP +  L+KYP+ + RV+D+K +L         DF+VITGSLYFIS++RQ
Sbjct: 394 HYPNAYPLEKYPERFGRVADFKDFLALRKHAKADDFFVITGSLYFISEIRQ 444
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1717

A DNA sequence (GBSx1821) was identified in *S.agalactiae* <SEQ ID 5343> which encodes the amino acid sequence <SEQ ID 5344>. This protein is predicted to be rarD. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.31    Transmembrane  130 - 146 ( 125 - 151)
   INTEGRAL    Likelihood =-10.24    Transmembrane  269 - 285 ( 262 - 291)
   INTEGRAL    Likelihood = -7.75    Transmembrane  212 - 228 ( 207 - 233)
   INTEGRAL    Likelihood = -5.52    Transmembrane   80 -  96 (  75 -  99)
   INTEGRAL    Likelihood = -4.14    Transmembrane  106 - 122 ( 104 - 125)
   INTEGRAL    Likelihood = -3.50    Transmembrane  182 - 198 ( 180 - 204)
   INTEGRAL    Likelihood = -2.44    Transmembrane   40 -  56 (  39 -  57)
   INTEGRAL    Likelihood = -0.96    Transmembrane  153 - 169 ( 152 - 169)
   INTEGRAL    Likelihood = -0.32    Transmembrane  251 - 267 ( 250 - 267)

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07585 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 109/288 (37%), Positives = 185/288 (63%), Gaps = 6/288 (2%)

Query: 7   GIILGLSAYVLWGLLSLYWKLLSGIEAYSTFAYRIIFTVLTMLIYMLVSGRKTVYLKDLK 66
           G+I  +SAY++WG L LYWKL+  + A    A+RI++++  M+I + V  +    ++++
Sbjct: 8   GVIAAISAYLIWGFLPLYWKLVDEVPASEMLAHRIVWSLGFMVILLAVMKKNRQVMREIL 67

Query: 67  GLVNNKKSFWTMFVASILISINWLVYIFAVTHGHATEASLGYYMMPIISILLSVLVLREH 126
             + NKK+ + + VA+ILIS+NW ++I+AV+     EASLGYY+ P+I++LL+++ LRE
Sbjct: 68  DTLANKKTAFGITVAAILISMNWFIFIYAVSSDKVIEASLGYYINPLINVLLAIVFLRES 127

Query: 127 LARVVSLAILIAIMGVGILVYQTGHFPLISLTLALSFGFYGLLKKSISLSSDFSMLVESS 186
           L++    + L+A  GV  +    G FP ++  LA+SFG YGL+KK +SLS+  S+ +E+
Sbjct: 128 LSKWEVASFLLAAAGVLNITLHYGSFPWVAFALAISFGVYGLIKKVVSLSAWASLTIETL 187

Query: 187 FIAPFALIYIVFF-----AKDFLTDYNILQLVLLSLSGIITAVPLLLFAEAIKRAPLNII 241
            + PFAL+++++      A  F  ++ +    L+  SG  TA+PLLLFA   KR   ++I
Sbjct: 188 IMTPFALLFLLYIPLSGGASAFSLNH-LSTAWLIIASGAATALPLLLFATGAKRISFSLI 246

Query: 242 GFIQYINPTIQLLLALFIFKETIVSGEVIGFIFIWLAILVFSIGQVHT 289
           GF+QY+ PTI L+L +F+F+E     + + F+ IW  +++F+I +  T
Sbjct: 247 GFLQYLAPTIMLMLGVFLFQEPFSRVQFVSFLLIWTGLIIFTISRSRT 294
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8885> and protein <SEQ ID 8886> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      5.30
GvH: Signal Score (-7.5): -1.64
     Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 9 value: -12.31 threshold:  0.0
   INTEGRAL    Likelihood =-12.31   Transmembrane  130 - 146 ( 125 - 151)
   INTEGRAL    Likelihood =-10.24   Transmembrane  269 - 285 ( 262 - 291)
   INTEGRAL    Likelihood = -7.75   Transmembrane  212 - 228 ( 207 - 233)
   INTEGRAL    Likelihood = -5.52   Transmembrane   80 -  96 (  75 -  99)
   INTEGRAL    Likelihood = -4.14   Transmembrane  106 - 122 ( 104 - 125)
   INTEGRAL    Likelihood = -3.50   Transmembrane  182 - 198 ( 180 - 204)
   INTEGRAL    Likelihood = -2.44   Transmembrane   40 -  56 (  39 -  57)
   INTEGRAL    Likelihood = -0.96   Transmembrane  153 - 169 ( 152 - 169)
   INTEGRAL    Likelihood = -0.32   Transmembrane  251 - 267 ( 250 - 267)
   PERIPHERAL  Likelihood =  7.96     229
 modified ALOM score:   2.96

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02052(319 - 1152 of 1485)
GP|9654601|gb|AAF93371.1||AE004110(13 - 289 of 302) rarD protein {Vibrio cholerae}
%Match = 20.4
%Identity = 37.7  %Similarity = 66.3
Matches = 104  Mismatches = 89  Conservative Sub.s = 79

117       147       177       207       237       267       297       327
KDIVNLW*RNLK**NKSALKMVRMLLICLEQDRR*WFCVRKKKNKQLSQS*VNYV*VDRFKCLILSEKE*ELRKDNLGII
                                                                             ||:
                                                                 MFMTPDQQDAKKGIL
                                                                         10

357       387       417       441       471       501       531       561
LGLSAYVLWGLLSLYWKLLSGIEAYSTFAYRII--FTVLTMLIYMLVSGRKTVYLKDLKGLVNNKKSFWTMFVASILISI
| :||| :||:  :|:| |  : |   :::|::  | :| :||::    |  |      |:|:  : || ::| ::|:
LAISAYTMWGIAPIYFKALGAVSALEILSHRVVWSFVLLAVLIHLGRRWSVV------GVVHTPRKFWLLLVTALLVGG
             30        40        50        60               70        80

591       621       651       681       711       741       771       801
NWLVYIFAVTHGHATEASLGYYMMPIISILLSVLVLREHLARVVSLAILIAIMGVGILVYQTGHFPLISLTLALSFGFYG
|||::|:::   |  :||||||: |::::|| :| | | | ::  :|: :| :|||| :   |  |:::: || ||||||
NWLIFIWSINANHMLDASLGYYINPLLNVLLGMLFLGERLRKLQWFAVALAAIGVGIQLVVFGSVPIVAIALATSFGFYG
        100       110       120       130       140       150       160

831       861       891       921       942       972       1002      1032
LLKKSISLSSDFSMLVESSFIAPFALIYIVFFAKDFLTDY--NILQL-VLLSLSGIITAVPLLLFAEAIKRAPLNIIGFI
||:| | : :   :::|: |: | | ||::::|    :|   |  || :||  :|::| :||| |  |  |  |: :||
LLRKKIQVDAQTGLFLETLFMLPAAAIYLIWLADTPTSDMALNTWQLNLLLVCAGVVTTLPLLCFTGAAARLKLSTLGFF
        180       190       200       210       220       230       240

1062      1092      1122      1152      1182      1212      1242      1272
QYINPTIQLLLALFIFKETIVSGEVIGFIFIWLAILVFSIGQVHTMLKKGK*DDLSRSARMDS**ISFWY*TRFGTYEMD
||| |:: :|||:::: |   | : | | ||| |:::||:
QYIGPSLMFLLAVLVYGEAFTSDKAITFAFIWSALVIFSVDGLKAGHAARRAR
        260       270       280       290       300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1718

A DNA sequence (GBSx1822) was identified in *S.agalactiae* <SEQ ID 5345> which encodes the amino acid sequence <SEQ ID 5346>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5200(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1719

A DNA sequence (GBSx1823) was identified in *S.agalactiae* <SEQ ID 5347> which encodes the amino acid sequence <SEQ ID 5348>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0881(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44297 GB:U41735 homoserine kinase homolog [Streptococcus pneumoniae]
 Identities = 188/289 (65%), Positives = 232/289 (80%), Gaps = 1/289 (0%)

Query: 1   MRIIVPATSANIGPGFDSIGVALSKYLIIEVLEESTEWLVEHNLVN-IPKDHTNLLIQTA 59
           M+IIVPATSANIGPGFDS+GVA++KYL IEV EE  EWL+EH +   IP D  NLL+  A
Sbjct: 1   MKIIVPATSANIGPGFDSVGVAVTKYLQIEVSEERDEWLIEHQIGKWIPHDERNLLLTIA 60

Query: 60  LHVKSDLAPHRLKMFSDIPLARGLGSSSSVIVAGIELANQLGNLALSQKEKLEIATRLEG 119
           L +  DL P RLKM SD+PLARGLGSSSSVIVAGIELANQLG L LS  EKL++AT++EG
Sbjct: 61  LQIVPDLQPRRLKMTSDVPLARGLGSSSSVIVAGIELANQLGQLNLSDHEKLQLATKIEG 120

Query: 120 HPDNVAPAIFGDLVISSIVKNDIKSLEVMFPDSSFIAFIPNYELKTSDSRNVLPQKLSYE 179
           HPDNVAPAI+G+LVI+S V+  + ++    FP+  F+A+IPNYEL+T DSR+VLP+KLSY+
Sbjct: 121 HPDNVAPAIYGNLVIASSVEGQVSAIVADFPECDFLAYIPNYELRTRDSRSVLPKKLSYK 180

Query: 180 DAVASSSVANVMVASLLKGDLVTAGWAIERDLFHERYRQPLVKEFEVIKQISTQNGAYAT 239
           +AVA+SS+ANV VA+LL GD+VTAG AIE DLFHERYRQ LV+EF +IKQ++ +NGAYAT
Sbjct: 181 EAVAASSIANVAVAALLAGDMVTAGQAIEGDLFHERYRQDLVREFAMIKQVTKENGAYAT 240

Query: 240 YLSGAGPTVMVLCSKEKEQAIVTELSKLCLGGQIQVLNIERKGVRVEKR 288
           YLSGAGPTVMVL S +K   I  EL K    G++  L ++ +GVRVE +
Sbjct: 241 YLSGAGPTVMVLASHDKMPTIKAELEKQPFKGKLHDLRVDTQGVRVEAK 289
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1720

A DNA sequence (GBSx1824) was identified in *S.agalactiae* <SEQ ID 5349> which encodes the amino acid sequence <SEQ ID 5350>. This protein is predicted to be homoserine dehydrogenase (hom). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9857> which encodes amino acid sequence <SEQ ID 9858> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA65713 GB:X96988 hom [Lactococcus lactis]
 Identities = 221/432 (51%), Positives = 307/432 (70%), Gaps = 11/432 (2%)

Query: 15  MTIKIALLGFGTVAKGIPYLLKENQHKLLSLEGEDIVIDKVLVRDNESRQRFINQGFTYN 74
           M + IA+LGFGTV  G+P LL EN+ KL  +  E+IVI KVL+RDN++ ++  +QGF Y+
Sbjct: 1   MAVNIAILGFGTVGTGLPTLLSENKEKLAKILDEEIVISKVLMRDNKAIEKARSQGFNYD 60

Query: 75  FVTEINTILQDSQIDIVVELMGGIEPAKTYLSQALGFGKHIVTANKDLIALHGKELMDLA 134
           FV  ++ IL DS+I IVVELMG IEPAKTY++QA+  GK++VTANKDL+A+HG EL  LA
Sbjct: 61  FVLNLDDILADSEISIVVELMGRIEPAKTYITQAIEAGKNVVTANKDLLAVHGVELRSLA 120

Query: 135 DARGLALFYEGAVAGGIPILRTLSHSFASDKMTRLLGILNGTSNFMLTKMFEEGWSYEQA 194
                +AL+YE AVAGGIPILRTL++SF+SDK+T LLGILNGTSNFM+TKM EEGW+Y+++
Sbjct: 121 QKHHVALYYEAAVAGGIPILRTLANSFSSDKITHLLGILNGTSNFMMTKMSEEGWTYDES 180

Query: 195 LKKAQELGYAESDPTNDVEGIDTAYKATILSQFGFGMPIDFDDVNYKGISSIRSEDVEVA 254
           L KAQELGYAESDPTNDV+GID +YK  ILS+F FGM +  DD+   G+ SI+  DVE+A
Sbjct: 181 LAKAQELGYAESDPTNDVDGIDASYKLAILSEFAFGMTLAPDDIAKSGLRSIQKTDVEIA 240

Query: 255 QEMGFAIKLVADLRETPTGISVDVSPTLISQKHPLAAVNHVMNAVFIESIGIGQSLFYGP 314
           Q+ G+ +KL  ++ E  +GI  +VSPT + + HPLA+VN VMNAVFIES GIG S+FYG
Sbjct: 241 QQFGYVLKLTGEINEVDSGIFAEVSPTFLPKSHPLASVNGVMNAVFIESEGIGDSVFYGA 300

Query: 315 GAGQNPTATSVLADIIDISRSIRSQIKIKPMNTYHCPCRLSMQSDIFNEYYLAISLRNAE 374
           GAGQ PTATSVLADI+ I + ++      K  N Y      L+   DI N+YY ++     E
Sbjct: 301 GAGQKPTATSVLADIVRIVKRVKDGTIGKSFNEYARSTSLANPHDIENKYYFSV-----E 355

Query: 375 DSDTLGR------YFEQENIGLKNVIEKALGDKQQEIYVLTDEVSQEKITQFIEEFPESG 428
             D+ G+       F  EN+  + V+++    K+  + +++ ++++ +++    ++   +
Sbjct: 356 TPDSTGQLLLLVELFTSENVSFEQVLQQKGNGKRAVVVIISHKINRVQLSAIQDKLNQEK 415

Query: 429 VIQLINVFKVIG 440
             +L+N FKV+G
Sbjct: 416 DFKLLNRFKVLG 427
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1721

A DNA sequence (GBSx1825) was identified in *S.agalactiae* <SEQ ID 5351> which encodes the amino acid sequence <SEQ ID 5352>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4548(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1722

A DNA sequence (GBSx1826) was identified in *S.agalactiae* <SEQ ID 5353> which encodes the amino acid sequence <SEQ ID 5354>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.79    Transmembrane    20 -   36 (   14 -   41)

----- Final Results -----
               bacterial membrane --- Certainty=0.6116(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15906 GB:Z99123 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 105/272 (38%), Positives = 149/272 (54%), Gaps = 20/272 (7%)

Query: 25  FLLIALIGIFLFFNNRSKQEIKT-----KTNASSHRKIVTSIKKKK-----WIKQKTPVK 74
            FL I L+G  L     + QE        K       K    ++KK+     WIK + P K
Sbjct: 5   FLSIFLLGSCLALAACADQEANAEQPMPKAEQKKPEKKAVQVQKKEDDTSAWIKTEKPAK 64

Query: 75  IPILMYHAVHVMDPSEAASANLIVAPDIFESHIKRLKKEGYYFLAPNEAYRALNENALPE 134
            +PILMYH++       ++  +L V    FE+H+K L   GY  L P EA   L ++  P
Sbjct: 65  LPILMYHSI-------SSGNSLRVPKKEFEAHMKWLHDNGYQTLTPKEASLMLTQDKKPS 117

Query: 135 KKVIWITFDDGNADFYTKAYPILKKYKVKATNNIITGFVQEGRESNLNVQQMLEMKQNGM 194
            +K + ITFDDG  D Y  AYP+LKKY +KAT  +I   +  G + +L  +QM EM Q+G+
Sbjct: 118 EKCVLITFDDGYTDNYQDAYPVLKKYGMKATIFMIGKSI--GHKHHLTEEQMKEMAQHGI 175

Query: 195 SFQGHTVTHPNLSLLTPELQTQEMTLSKQFLDQKLSQDTLAIAYPSGRYNPTTLDIASQY 254
            S + HT+ H  L+ LTP+ Q  EM  SK+  D    Q T  I+YP GRYN  TL  A +
Sbjct: 176 SIESHTIDHLELNGLTPQQQQSEMADSKKLFDNMFHQQTTIISYPVGRYNEETLKAAEKT 235

Query: 255 -YKLGLTTNEGVATKDNGLLSLNRIRILPTTS 285
             Y++G+TT  G A++D G+ +L+R+R+ P  S
Sbjct: 236 GYQMGVTTEPGAASRDQGMYALHRVRVSPGMS 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5355> which encodes the amino acid sequence <SEQ ID 5356>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
```

```
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15906 GB:Z99123 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 97/240 (40%), Positives = 140/240 (57%), Gaps = 9/240 (3%)

Query: 71  KKTHFDSSKSQKKAHSKLTWTKQETPVKIPILMYHAIHVMSPEETANANLIVNPDLFDQQ 130
           KK    + + QKK      W K E P K+PILMYH+I       ++  +L V    F+
Sbjct: 37  KKPEKKAVQVQKKEDDTSAWIKTEKPAKLPILMYHSI-------SSGNSLRVPKKEFEAH 89

Query: 131 LQKMKDEGYYFLSPEEVYRALSNNELPAKKVVWLTFDDSMIDFYNVAYPILKKYDAKATN 190
           ++ + D GY  L+P+E    L+ ++ P++K V +TFDD   D Y  AYP+LKKY  KAT
Sbjct: 90  MKWLHDNGYQTLTPKEASLMLTQDKKPSEKCVLITFDDGYTDNYQDAYPVLKKYGMKATI 149

Query: 191 NVITGLTEMGSAANLTLKQMKEMKQVGMSFQDHTVNHPDLEQASPDVQTTEMKDSKDYLD 250
            +I     +G   +LT +QMKEM Q G+S + HT++H +L   +P  Q +EM DSK   D
Sbjct: 150 FMIG--KSIGHKHHLTEEQMKEMAQHGISIESHTIDHLELNGLTPQQQQSEMADSKKLFD 207

Query: 251 KQLNQNTIAIAYPSGRYNDTTLQIAARLNYKLGVTTNEGIASAANGLLSLNRIRILPNMS 310
              +Q T  I+YP GRYN+ TL+ A +  Y++GVTT  G AS   G+ +L+R+R+ P MS
Sbjct: 208 NMFHQQTTIISYPVGRYNEETLKAAEKTGYQMGVTTEPGAASRDQGMYALHRVRVSPGMS 267
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/265 (57%), Positives = 199/265 (74%), Gaps = 4/265 (1%)

Query: 33  IFLFFNNRSKQEIKTK---TNASSHRKIVTSIKKKKWIKQKTPVKIPILMYHAVHVMDPS 89
           I LF + ++ ++  TK   T+  S +    +  K  W KQ+TPVKIPILMYHA+HVM P
Sbjct: 54  ISLFHHKKTAKKETTKLKKTHFDSSKSQKKAHSKLTWTKQETPVKIPILMYHAIHVMSPE 113

Query: 90  EAASANLIVAPDIFESHIKRLKKEGYYFLAPNEAYRALNENALPEKKVIWITFDDGNADF 149
           E A+ANLIV PD+F+  ++++K EGYYFL+P E YRAL+ N LP KKV+W+TFDD   DF
Sbjct: 114 ETANANLIVNPDLFDQQLQKMKDEGYYFLSPEEVYRALSNNELPAKKVVWLTFDDSMIDF 173

Query: 150 YTKAYPILKKYKVKATNNIITGFVQEGRESNLNVQQMLEMKQNGMSFQGHTVTHPNLSLL 209
           Y  AYPILKKY  KATNN+ITG  + G  +NL ++QM EMKQ GMSFQ HTV HP+L
Sbjct: 174 YNVAYPILKKYDAKATNNVITGLTEMGSAANLTLKQMKEMKQVGMSFQDHTVNHPDLEQA 233

Query: 210 TPELQTQEMTLSKQFLDQKLSQDTLAIAYPSGRYNPTTLDIASQY-YKLGLTTNEGVATK 268
           +P++QT EM  SK +LD++L+Q+T+AIAYPSGRYN TTL IA++  YKLG+TTNEG+A+
Sbjct: 234 SPDVQTTEMKDSKDYLDKQLNQNTIAIAYPSGRYNDTTLQIAARLNYKLGVTTNEGIASA 293

Query: 269 DNGLLSLNRIRILPTTSDDDLIKTI 293
            NGLLSLNRIRILP  S ++L++T+
Sbjct: 294 ANGLLSLNRIRILPNMSPENLLQTM 318
```

SEQ ID 5354 (GBS287d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 3 & 4; MW 57kDa) and in Figure 185 (lane 2; MW 57kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 6; MW 32kDa) and in Figure 181 (lane 5; MW 32kDa).

Purified GBS287d-GST is shown in Figure 243, lanes 10-11; purified GBS287d-His is shown in Figure 234, lanes 7-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1723

A DNA sequence (GBSx1828) was identified in *S.agalactiae* <SEQ ID 5357> which encodes the amino acid sequence <SEQ ID 5358>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1724

A DNA sequence (GBSx1829) was identified in *S.agalactiae* <SEQ ID 5359> which encodes the amino acid sequence <SEQ ID 5360>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3352(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1725

A DNA sequence (GBSx1830) was identified in *S.agalactiae* <SEQ ID 5361> which encodes the amino acid sequence <SEQ ID 5362>. This protein is predicted to be glycine betaine transporter BetL (opuD). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.68   Transmembrane  439 - 455 ( 435 - 491)
   INTEGRAL    Likelihood =-12.10   Transmembrane  256 - 272 ( 249 - 281)
   INTEGRAL    Likelihood =-11.30   Transmembrane  464 - 480 ( 456 - 491)
   INTEGRAL    Likelihood =-10.83   Transmembrane   49 -  65 (  44 -  74)
   INTEGRAL    Likelihood =-10.40   Transmembrane   11 -  27 (   5 -  34)
   INTEGRAL    Likelihood = -9.98   Transmembrane  396 - 412 ( 390 - 419)
   INTEGRAL    Likelihood = -9.29   Transmembrane  224 - 240 ( 220 - 247)
   INTEGRAL    Likelihood = -7.11   Transmembrane  347 - 363 ( 341 - 366)
   INTEGRAL    Likelihood = -2.87   Transmembrane  143 - 159 ( 143 - 159)
   INTEGRAL    Likelihood = -2.60   Transmembrane  192 - 208 ( 191 - 208)
   INTEGRAL    Likelihood = -1.44   Transmembrane   86 - 102 (  86 - 105)

----- Final Results -----
               bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
```

```
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD30266 GB:AF102174 glycine betaine transporter BetL [Listeria
          monocytogenes]
 Identities = 277/503 (55%), Positives = 365/503 (72%), Gaps = 1/503 (0%)

Query: 4   KHITPVFTGSLIVSLILVLLGIIVPRGFQSWTQILREQVSTNFGWLYLLLVTSILALCVF 63
            K +T VF GS  + L+ VL G  +P  F+++T  +++ +++NFGW YL++V  I+  C+F
Sbjct: 2   KKLTNVFWGSGFLVLLAVLFGAFLPEQFETFTNHIQKFLTSNFGWYYLIVVAIIIIFCLF 61

Query: 64  FIMSPLGQIRLGQPHSRPEYSTVSWIAMMFSAGMGIGLVFYGAAEPLSHFAISTPGAPKE 123
             ++SP+G IRLG+P   P YS  SW AM+FSAGMGIGLVF+GAAEPLSH+A+  PG
Sbjct: 62  LVLSPIGSIRLGKPGEEPGYSNKSWFAMLFSAGMGIGLVFWGAAEPLSHYAVQAPGGEVG 121

Query: 124 SQTALADAFRFTFFHWGIHAWAVYALVALALAYFGFRKQEKYLLSVTLKPLFGDKTDGWL 183
            +Q A+ DA R++FFHWGI AW++YA+VALALAYF FRK    L+S TL P+ G    G +
Sbjct: 122 TQAAMKDALRYSFFHWGISAWSIYAIVALALAYFKFRKNAPGLISATLYPILGKHAKGPI 181

Query: 184 GKIVDITTVVATVIGVATTLGFGAAQINGGLSFLLGVPNNAFVQIVIILITTALFVMSAL 243
            G+++DI  V ATVIGVATTLG GA QINGGL++L GVPNN  VQ  II+I T LF++SA+
Sbjct: 182 GQLIDIIAVFATVIGVATTLGLGAQQINGGLTYLFGVPNNFTVQFTIIVIVTILFMLSAM 241

Query: 244 SGLGKGVKILSNLNLILAVALLALVIVLGPTVRIFDTLTESLGSYLQNFFGMSFRAAAFD 303
            SGL KG+++LSN+N+ +A  LL L ++LGPT+ I +  T S G YLQN   MSF+ A
Sbjct: 242 SGLDKGIQLLSNVNIYVAGVLLVLTLILGPTLFIMNNFTNSFGDYLQNIIQMSFQTAPDA 301

Query: 304 NTKRSWIDNWTIFYWAWWISWSPFVGVFIARISKGRSIREFLTVVLLIPTLLSFVWFAAF 363
               R WID+WTIFYWAWW+SWSPFVG+FIARIS+GR+IR+FL  V+++P L+S  WFA F
Sbjct: 302 PDARKWIDSWTIFYWAWWLSWSPFVGIFIARISRGRTIRQFLLGVIVLPALVSVFWFAVF 361

Query: 364 GTLSTQVQQLG-TNLTKFATEEVLFATFNHYTLGWLLSIIAIILIFSFFITSADSATYVL 422
            G  +  V+Q G + L+  ATE+VLF  FN +  G +LSI+A+ILI  FFITSADSAT+VL
Sbjct: 362 GGSAIFVEQHGNSGLSSLATEQVLFGVFNEFPGGMMLSIVAMILIAVFFITSADSATFVL 421

Query: 423 AMLTEDGNLNPKNRTKVIWGLVLAVIAIVLLLSGGLLALQNVLIIVALPFSFVMILMMLA 482
             M T  G+LNP N  KV WGL+ A IA VLL +GGL ALQN  II A PFS V+ILM+++
Sbjct: 422 GMQTTGGSLNPPNSVKVTWGLLQAGIASVLLYAGGLTALQNASIIAAFPFSIVIILMIVS 481

Query: 483 LLVELFHEKKEMGLSISPDRYPR 505
            L V L  E++++GL + P +  R
Sbjct: 482 LFVSLTREQEKLGLYVRPKKSQR 504
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8887> and protein <SEQ ID 8888> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      15.28
GvH: Signal Score (-7.5): -4.24
     Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 11 value: -12.68 threshold:  0.0
   INTEGRAL    Likelihood =-12.68   Transmembrane  439 - 455 ( 435 - 491)
   INTEGRAL    Likelihood =-12.10   Transmembrane  256 - 272 ( 249 - 281)
   INTEGRAL    Likelihood =-11.30   Transmembrane  464 - 480 ( 456 - 491)
   INTEGRAL    Likelihood =-10.83   Transmembrane   49 -  65 (  44 -  74)
   INTEGRAL    Likelihood =-10.40   Transmembrane   11 -  27 (   5 -  34)
   INTEGRAL    Likelihood = -9.98   Transmembrane  396 - 412 ( 390 - 419)
   INTEGRAL    Likelihood = -9.29   Transmembrane  224 - 240 ( 220 - 247)
   INTEGRAL    Likelihood = -7.11   Transmembrane  347 - 363 ( 341 - 366)
   INTEGRAL    Likelihood = -2.87   Transmembrane  143 - 159 ( 143 - 159)
   INTEGRAL    Likelihood = -2.60   Transmembrane  192 - 208 ( 191 - 208)
   INTEGRAL    Likelihood = -1.44   Transmembrane   86 - 102 (  86 - 105)
```

```
   PERIPHERAL  Likelihood =  3.50      319
 modified ALOM score:   3.04

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02057(310 - 1821 of 2145)
GP|4835822|gb|AAD30266.1|AF102174_1|AF102174(2 - 506 of 507) glycine betaine transporter
BetL {Listeria monocytogenes} PIR|T48645|T48645 glycine betaine transport protein betL
[validated] - Listeria monocytogenes
%Match = 38.7
%Identity = 54.9  %Similarity = 74.7
Matches = 277  Mismatches = 127  Conservative Sub.s = 100

54        84        114       144       174       204       234       264
IQGGHHYRNYRLEVLKIQDMVVS*ANLDLMPLSTNIWYLHQIVINH*VKHKNQIMLFGSFLLRRQGEVLIQVVKMRGVFI

294       324       354       384       414       444       474       504
KVCYTILV*EEILSKKHITPVFTGSLIVSLILVLLGIIVPRGFQSWTQILREQVSTNFGWLYLLLVTSILALCVFFIMSP
               | :| || ||  : |: ||:|  :|  |:::|  ::: :::|||| ||::|  |: :|:|:::||
              MKKLTNVFWGSGFLVLLAVLFGAFLPEQFETFTNHIQKFLTSNFGWYYLIVVAIIIIFCLFLVLSP
                      10        20        30        40        50        60

534       564       594       624       654       684       714       744
LGQIRLGQPHSRPEYSTVSWIAMMFSAGMGIGLVFYGAAEPLSHFAISTPGAPKESQTALADAFRFTFFHWGIHAWAVYA
:| ||||:|   | ||  || ||:||||||||||:||||||||:|:  ||    :| |: ||:|::|||||| ||::||
IGSIRLGKPGEEPGYSNKSWFAMLFSAGMGIGLVFWGAAEPLSHYAVQAPGGEVGTQAAMKDALRYSFFHWGISAWSIYA
            80        90       100       110       120       130       140

774       804       834       864       894       924       954       984
LVALALAYFGFRKQEKYLLSVTLKPLFGDKTDGWLGKIVDITTVVATVIGVATTLGFGAAQINGGLSFLLGVPNNAFVQI
:|||||||| |||    |:| || |::|    | :|:::||  | ||||||||||||:|| ||||||::|:|||||  ||
IVALALAYFKFRKNAPGLISATLYPILGKHAKGPIGQLIDIIAVFATVIGVATTLGLGAQQINGGLTYLFGVPNNFTVQF
           160       170       180       190       200       210       220

1014      1044      1074      1104      1134      1164      1194      1224
VIILITTALFVMSALSGLGKGVKILSNLNLILAVALLALVIVLGPTVRIFDTLTESLGSYLQNFFGMSFRAAAFDNTKRS
 ||:| | ||::||:||| ||:::|||:|: :|  || | ::||||: | : :| |:| ||||   |||: |      |
TIIVIVTILFMLSAMSGLDKGIQLLSNVNIYVAGVLLVLTLILGPTLFIMNNFTNSFGDYLQNIIQMSFQTAPDAPDARK
           240       250       260       270       280       290       300

1254      1284      1314      1344      1374      1404      1431      1461
WIDNWTIFYWAWWISWSPFVGVFIARISKGRSIREFLTVVLLIPTLLSFVWFAAFGTLSTQVQQLGTN-LTKFATEEVLF
|||:|||||||||:|||||||:||||||:||:||:||  |:::| |:|  ||| ||  :  |:| | : |: :|||:|||
WIDSWTIFYWAWWLSWSPFVGIFIARISRGRTIRQFLLGVIVLPALVSVFWFAVFGGSAIFVEQHGNSGLSSLATEQVLF
           320       330       340       350       360       370       380

1491      1521      1551      1581      1611      1641      1671      1701
ATFNHYTLGWLLSIIAIILIFSFFITSADSATYVLAMLTEDGNLNPKNRTKVIWGLVLAVIAIVLLLSGGLLALQNVLII
  || :  | :|||:|:|||  ||||||||||:|| | |  |:||| |  || |||: | || ||| :||| ||||  ||
GVFNEFPGGMMLSIVAMILIAVFFITSADSATFVLGMQTTGGSLNPPNSVKVTWGLLQAGIASVLLYAGGLTALQNASII
           400       410       420       430       440       450       460

1731      1761      1791      1821      1851      1881      1911      1941
VALPFSFVMILMMLALLVELFHEKKEMGLSISPDRYPRKNEPFKSYEE*KEARRLLFIG*SS*SDHHR**LVRYEFD*EK
 |:||| |:|||:::|:| |  |::::|| : | :  |
AAFPFSIVIILMIVSLFVSLTREQEKLGLYVRPKKSQRSQL
           480       490       500
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1726

A DNA sequence (GBSx1831) was identified in *S.agalactiae* <SEQ ID 5363> which encodes the amino acid sequence <SEQ ID 5364>. This protein is predicted to be succinic semialdehyde dehydrogenase (gabD-1). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2733(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9859> which encodes amino acid sequence <SEQ ID 9860> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD19405 GB:AF102543 succinic semialdehyde dehydrogenase
          [Zymomonas mobilis]
 Identities = 229/455 (50%), Positives = 305/455 (66%), Gaps = 5/455 (1%)

Query: 10  MAYKTIYPYTNEVLHEFDNISDSDLEQSLDIAHALYKTWRKEDNVEERQNQLHKVADLLR 69
            MAY+++ P T E + ++ + SD  ++ S+D A  ++K    +  + ER   LHK A++ R
Sbjct: 1   MAYESVNPATGETVKKYPDFSDKQVKDSVDRAATVFKNDWSQRTIAERSKVLHKAAEIFR 60

Query: 70  KDRDKYAEVMTKDMGKLFTEAQGEVDLCADIADYYADNGQKFLKPVPLESPNGEAYYLKQ 129
             D DKYA+++T DMGK   EA+GEV+L ADI DYYA NG+KFL P  +E   G A
Sbjct: 61  SDVDKYAKLLTIDMGKKIAEARGEVNLSADILDYYAKNGEKFLAPQKVEEKPG-AVVKAF 119

Query: 130 AVGVLLAVEPWNFPFYQIMRVFAPNFIVGNTMLLKHASICPASAQAFEDLVREAGAPEGA 189
             +G+LLA+EPWNFP+YQ+ R+  P  I GN +L+KH+S  P SA AFE ++ EAGAP+G
Sbjct: 120 PLGLLLAIEPWNFPYYQLARIAGPYLIAGNALLVKHSSSVPQSAHAFEAVLEEAGAPKGI 179

Query: 190 FKNIFASYDQVSNLISDPRVAGVCLTGSERGGASIAAEAGKNLKKSSMELGGNDAFLILD 249
            + N+ AS DQVS +I DPRV GV +TGS   GA +AA+AGK  KKS MELGG+DAF++LD
Sbjct: 180 YTNLDASPDQVSQIIEDPRVRGVTVTGSASVGAELAAKAGKMWKKSVMELGGSDAFIVLD 239

Query: 250 DADFD--LLSKTIFFARLYNAGQVCTSSKRFIVMADKYDE-FVNMVVETFKSAKWGDPMD 306
              D D  L+ K  +  RL+NAGQV  ++KRFI++  K  E F   + + F++ K GDPMD
Sbjct: 240 GVDIDDKLIDKAAY-GRLFNAGQVFCAAKRFIIVGQKRAELFTEKLKQRFEALKIGDPMD 298

Query: 307 SETTLAPLSSAGAKDDVLKQIKLAVDHGAEVVFGNDTIDHPGNFVMPTVLTNITKANPIY 366
              T L PLSS GA+D V+KQ++ AV +GA++V G   I+  G F+   +LT+I + NP Y
Sbjct: 299 ESTDLGPLSSVGARDQVVKQVEKAVQNGAKLVCGGKAIEGKGAFMKAGILTDIKRENPAY 358

Query: 367 NQEIFGPVASIYKVDTEEEAIALANDSSYGLGSTVFSSDPEHAKKVAAQIETGMTFINSG 426
             +E FGP+A IY V  E EAI LANDS YGLG  VF+ D E  +KVA QIETGM  IN
Sbjct: 359 FEEFFGPIAQIYAVKDEAEAIELANDSPYGLGGAVFAPDVEQGRKVAEQIETGMVAINKP 418

Query: 427 WTSLPELPFGGIKNSGYGRELSQLGFDAFVNEHLV 461
              + PELPFGG+K+SGYGRELS  G   F+N  L+
Sbjct: 419 LWTAPELPFGGVKHSGYGRELSHFGIQEFINWKLI 453
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5365> which encodes the amino acid sequence <SEQ ID 5366>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2887(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 335/457 (73%), Positives = 397/457 (86%)

Query: 9   IMAYKTIYPYTNEVLHEFDNISDSDLEQSLDIAHALYKTWRKEDNVEERQNQLHKVADLL 68
           +MAY+TIYPYTNEVLH FDN++D  L   L+ AH LYK WRKED++EER+ QLH+VA++L
Sbjct: 1   VMAYQTIYPYTNEVLHTFDNMTDQGLADVLERAHLLYKKWRKEDHLEERKAQLHQVANIL 60

Query: 69  RKDRDKYAEVMTKDMGKLFTEAQGEVDLCADIADYYADNGQKFLKPVPLESPNGEAYYLK 128
           R+DRDKYAE+MTKDMGKLFTEAQGEV+LCADIADYYAD   +FL   PLE+ +G+AYYLK
Sbjct: 61  RRDRDKYAEIMTKDMGKLFTEAQGEVNLCADIADYYADKADEFLMSTPLETDSGQAYYLK 120

Query: 129 QAVGVLLAVEPWNFPFYQIMRVFAPNFIVGNTMLLKHASICPASAQAFEDLVREAGAPEG 188
           Q+ GV+LAVEPWNFP+YQIMRVFAPNFIVGN M+LKHASICP SAQ+FE+LV EAGA  G
Sbjct: 121 QSTGVILAVEPWNFPYYQIMRVFAPNFIVGNPMVLKHASICPRSAQSFEELVLEAGAEAG 180

Query: 189 AFKNIFASYDQVSNLISDPRVAGVCLTGSERGGASIAAEAGKNLKKSSMELGGNDAFLIL 248
           +  N+F SYDQVS +I+D RV GVCLTGSERGGASIA EAGKNLKK+++ELGG+DAF+IL
Sbjct: 181 SITNLFISYDQVSQVIADKRVVGVCLTGSERGGASIAEEAGKNLKKTTLELGGDDAFIIL 240

Query: 249 DDADFDLLSKTIFFARLYNAGQVCTSSKRFIVMADKYDEFVNMVVETFKSAKWGDPMDSE 308
           DDAD+D L K ++F+RLYNAGQVCTSSKRFIV+   YD F  ++ + FK+AKWGDPMD E
Sbjct: 241 DDADWDQLEKVLYFSRLYNAGQVCTSSKRFIVLDKDYDRFKELLTKVFKTAKWGDPMDPE 300

Query: 309 TTLAPLSSAGAKDDVLKQIKLAVDHGAEVVFGNDTIDHPGNFVMPTVLTNITKANPIYNQ 368
           TTLAPLSSA AK DVL QIKLA+DHGAE+V+G + IDHPG+FVMPT++  +TK NPIY Q
Sbjct: 301 TTLAPLSSAQAKADVLDQIKLALDHGAELVYGGEAIDHPGHFVMPTIIAGLTKDNPIYYQ 360

Query: 369 EIFGPVASIYKVDTEEEAIALANDSSYGLGSTVFSSDPEHAKKVAAQIETGMTFINSGWT 428
           EIFGPV  IYKV +EEEAI +ANDS+YGLG T+FSS+ EHAK VAA+IETGM+FINSGWT
Sbjct: 361 EIFGPVGEIYKVSSEEEAIEVANDSNYGLGGTIFSSNQEHAKAVAAKIETGMSFINSGWT 420

Query: 429 SLPELPFGGIKNSGYGRELSQLGFDAFVNEHLVFTPN 465
           SLPELPFGGIK+SGYGRELS+LGF +FVNEHL++ PN
Sbjct: 421 SLPELPFGGIKHSGYGRELSELGFTSFVNEHLIYIPN 457
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1727

A DNA sequence (GBSx1832) was identified in *S.agalactiae* <SEQ ID 5367> which encodes the amino acid sequence <SEQ ID 5368>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1728

A DNA sequence (GBSx1833) was identified in *S.agalactiae* <SEQ ID 5369> which encodes the amino acid sequence <SEQ ID 5370>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.91    Transmembrane   94 - 110 (  86 - 115)
   INTEGRAL    Likelihood = -7.75    Transmembrane  154 - 170 ( 150 - 176)
   INTEGRAL    Likelihood = -7.11    Transmembrane  316 - 332 ( 312 - 339)
   INTEGRAL    Likelihood = -6.16    Transmembrane  258 - 274 ( 253 - 278)
   INTEGRAL    Likelihood = -2.71    Transmembrane  218 - 234 ( 217 - 234)
   INTEGRAL    Likelihood = -1.49    Transmembrane  286 - 302 ( 283 - 302)
   INTEGRAL    Likelihood = -0.96    Transmembrane   73 -  89 (  73 -  89)
   INTEGRAL    Likelihood = -0.27    Transmembrane  121 - 137 ( 121 - 137)

----- Final Results -----
               bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9861> which encodes amino acid sequence <SEQ ID 9862> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75219 GB:AE000305 orf, hypothetical protein [Escherichia coli K12]
 Identities = 102/331 (30%), Positives = 172/331 (51%), Gaps = 26/331 (7%)

Query: 12  IPGLILCFIIA-IPSWLLGLYLPLIGAPVF-----AILIGIIVGSFYQNR--QLFNKGIA 63
            IPGL L  +I  +  W  G  +P +    F     AIL+G+++G+       +  + G+
Sbjct: 17  IPGLALSAVITGVALW--GGSIPAVAGAGFSALTLAILLGMVLGNTIYPHIWKSCDGGVL 74

Query: 64  FTSKYILQTAVVLLGFGLNLMQVMKVGISSLPIIIMTISISLIIAYVL-QKLFKLDKTIA 122
            F  +Y+L+  ++L GF L   Q+  VGIS + I ++T+S + ++A  L QK+F LDK  +
Sbjct: 75  FAKQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSSTFLLACFLGQKVFGLDKHTS 134

Query: 123 TLIGVGSSICGGSAIAATAPVINAKDDEVAQAISVIFLFNILAALIFPTLGNFIG--LSD 180
             LIG GSSICG +A+ AT PV+ A+  +V  A++ + +F  +A  ++P +   +    S
Sbjct: 135 WLIGAGSSICGAAAVLATEPVVKAEASKVTVAVATVVIFGTVAIFLYPAIYPLMSQWFSP 194

Query: 181 HGFALFAGTAVNDTSSVTAT--ATAWDAINHSNTLGGATIVKLTRTLAIIPITIVLSIYH 238
              F ++ G+ V++ + V A   A + DA N       A I K+ R + + P  I+L+
Sbjct: 195 ETFGIYIGSTVHEVAQVVAAGHAISPDAEN------AAVISKMLRVMMLAPFLILLAA-R 247

Query: 239 MKQTQKEQSVSVTKI-FPKFVLYFILASLLTTIVASLGFSLRIFEPLKVLSKFFIVMAMG 297
            +KQ     S   +KI  P F + FI+ ++  +          +   L  L  F + MAM
Sbjct: 248 VKQLSGANSGEKSKITIPWFAILFIVVAIFNSFHL---LPQSVVNMLVTLDTFLLAMAMA 304

Query: 298 AIGINTNVSKLIKTGGKSILLGAACWLGIII 328
            A+G+ T+VS L K G K +L+    +I+
Sbjct: 305 ALGLTTHVSALKKAGAKPLLMALVLFAWLIV 335
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5371> which encodes the amino acid sequence <SEQ ID 5372>. Analysis of this protein sequence reveals the following:

```
     Possible site: 37
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.29    Transmembrane   30 -  46 (  22 -  50)
   INTEGRAL    Likelihood = -8.12    Transmembrane  314 - 330 ( 311 - 338)
   INTEGRAL    Likelihood = -6.05    Transmembrane    8 -  24 (   7 -  29)
   INTEGRAL    Likelihood = -6.00    Transmembrane  150 - 166 ( 146 - 172)
   INTEGRAL    Likelihood = -5.57    Transmembrane  257 - 273 ( 252 - 277)
   INTEGRAL    Likelihood = -3.50    Transmembrane   91 - 107 (  87 - 108)
   INTEGRAL    Likelihood = -2.60    Transmembrane   69 -  85 (  68 -  87)
   INTEGRAL    Likelihood = -2.55    Transmembrane  289 - 305 ( 289 - 305)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC75219 GB:AE000305 orf, hypothetical protein [Escherichia
           coli]
 Identities = 100/329 (30%), Positives = 173/329 (52%), Gaps = 21/329 (6%)

Query: 8   LPGLLLCLLLALPAWCLGRLFPIIGAP----VFAILLGMLLA-LFYEHRDKTKEG-ISFT 61
           +PGL L  ++   A   G +  + GA       AILLGM+L    Y H  K+ +G + F
Sbjct: 17  IPGLALSAVITGVALWGGSIPAVAGAGFSALTLAILLGMVLGNTIYPHIWKSCDGGVLFA 76

Query: 62  SKYILQTAVVLLGFGLNLTQVMAVGMQSLPIIISTIATALLVAYGL-QKWLRLDVNTATL 120
            +Y+L+  ++L GF L  +Q+  VG+  + I + T+++  L+A  L QK   LD +T+ L
Sbjct: 77  KQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSSTFLLACFLGQKVFGLDKHTSWL 136

Query: 121 VGVGSSICGGSAVAATAPVIKAKDDEVAKAISVIFLFNMLAALLFPSLGQLLG--LSNEG 178
           +G GSSICG +AV AT PV+KA+  +V  A++ + +F  +A  L+P++  L+    S E
Sbjct: 137 IGAGSSICGAAAVLATEPVVKAEASKVTVAVATVVIFGTVAIFLYPAIYPLMSQWFSPET 196

Query: 179 FAIFAGTAVNDTSSVTATATAWDALHHSNTLDGATIVKLTRTLAILPITLGLSLYRAKKE 238
           F I+ G+ V++ + V A   A       +  + A I K+ R + + P  + L+  R K+
Sbjct: 197 FGIYIGSTVHEVAQVVAAGHAIS----PDAENAAVISKMLRVMMLAPFLILLAA-RVKQL 251

Query: 239 HDIVTEENFSLRKSFPRFILFFLLASLITTLMTSLGVSADSFHYLKTLSKFFIVMAMAAI 298
               + E   +  + P F + F++ ++  +      +     + L TL  F + MAMAA+
Sbjct: 252 SGANSGEKSKI--TIPWFAILFIVVAIFNSFHL---LPQSVVNMLVTLDTFLLAMAMAAL 306

Query: 299 GLNTNLVKLIKTGGQAILLGAI--CWVAI 325
           GL T++  L K G + +L+  +   W+ +
Sbjct: 307 GLTTHVSALKKAGAKPLLMALVLFAWLIV 335
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 225/333 (67%), Positives = 277/333 (82%), Gaps = 3/333 (0%)

Query: 11  KIPGLILCFIIAIPSWLLGLYLPLIGAPVFAILIGIIVGSFYQNRQLFNKGIAFTSKYIL 70
           K+PGL+LC ++A+P+W LG   P+IGAPVFAIL+G+++  FY++R    +GI+FTSKYIL
Sbjct: 7   KLPGLLLCLLLALPAWCLGRLFPIIGAPVFAILLGMLLALFYEHRDKTKEGISFTSKYIL 66

Query: 71  QTAVVLLGFGLNLMQVMKVGISSLPIIIMTISISLIIAYVLQKLFKLDKTIATLIGVGSS 130
           QTAVVLLGFGLNL QVM VG+ SLPIII TI+ +L++AY LQK  +LD   ATL+GVGSS
Sbjct: 67  QTAVVLLGFGLNLTQVMAVGMQSLPIIISTIATALLVAYGLQKWLRLDVNTATLVGVGSS 126

Query: 131 ICGGSAIAATAPVINAKDDEVAQAISVIFLFNILAALIFPTLGNFIGLSDHGFALFAGTA 190
           ICGGSA+AATAPVI AKDDEVA+AISVIFLFN+LAAL+FP+LG  +GLS+ GFA+FAGTA
Sbjct: 127 ICGGSAVAATAPVIKAKDDEVAKAISVIFLFNMLAALLFPSLGQLLGLSNEGFAIFAGTA 186

Query: 191 VNDTSSVTATATAWDAINHSNTLGGATIVKLTRTLAIIPITIVLSIYHMKQTQ---KEQS 247
           VNDTSSVTATATAWDA++HSNTL GATIVKLTRTLAI+PIT+ LS+Y  K+      E++
Sbjct: 187 VNDTSSVTATATAWDALHHSNTLDGATIVKLTRTLAILPITLGLSLYRAKKEHDIVTEEN 246

Query: 248 VSVTKIFPKFVLYFILASLLTTIVASLGFSLRIFEPLKVLSKFFIVMAMGAIGINTNVSK 307
            S+ K FP+F+L+F+LASL+TT++ SLG S   F  LK LSKFFIVMAM AIG+NTN+ K
Sbjct: 247 FSLRKSFPRFILFFLLASLITTLMTSLGVSADSFHYLKTLSKFFIVMAMAAIGLNTNLVK 306

Query: 308 LIKTGGKSILLGAACWLGIIIVSLTMQAILGTW 340
           LIKTGG++ILLGA CW+ I +VSL MQ  LG W
Sbjct: 307 LIKTGGQAILLGAICWVAITLVSLAMQLSLGIW 339
```

A related GBS gene <SEQ ID 8889> and protein <SEQ ID 8890> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     22.17
GvH: Signal Score (-7.5): -0.429999
     Possible site: 41
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 8 value:  -7.91 threshold:  0.0
   INTEGRAL    Likelihood = -7.91  Transmembrane   94 - 110 (  86 - 115)
   INTEGRAL    Likelihood = -7.75  Transmembrane  154 - 170 ( 150 - 176)
   INTEGRAL    Likelihood = -7.11  Transmembrane  316 - 332 ( 312 - 339)
   INTEGRAL    Likelihood = -6.16  Transmembrane  258 - 274 ( 253 - 278)
   INTEGRAL    Likelihood = -2.71  Transmembrane  218 - 234 ( 217 - 234)
   INTEGRAL    Likelihood = -1.49  Transmembrane  286 - 302 ( 283 - 302)
   INTEGRAL    Likelihood = -0.96  Transmembrane   73 -  89 (  73 -  89)
   INTEGRAL    Likelihood = -0.27  Transmembrane  121 - 137 ( 121 - 137)
   PERIPHERAL  Likelihood =  3.29     175
 modified ALOM score:   2.08

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02059(334 - 1284 of 1620)
EGAD|10465|EC2158(17 - 335 of 349) hypothetical 36.9 kd protein in lysp-nfo intergenic
region {Escherichia coli} OMNI|NT01EC2574 conserved hypothetical protein
SP|P33019|YEIH_ECOLI HYPOTHETICAL 36.9 KDA PROTEIN IN LYSP-NFO INTERGENIC REGION.
GP|405879|gb|AAA60511.1||U00007 yeiH {Escherichia coli} GP|1788482|gb|AAC75219.1||AE000305
orf, hypothetical protein {Escherichia coli} PIR|E64984|E64984 hypothetical 36.9 kD protein
in lysP-nfo intergenic region - Escherichia coli (strain K-12)
%Match = 12.7
%Identity = 32.3  %Similarity = 57.1
Matches = 103  Mismatches = 125  Conservative Sub.s = 79

270       300       330       360       390                 435       462
YSGPLSVFLSRFKACDIIVNVRRTIMLFKEKIPGLILCFIIAIPSWLLGLYLPLI-----GAPVFAILIGIIVG-SFYQN
                               |||| |  :|     | |  :| :      |  :|||:|:::| : | :
               MTNITLQKQHRTLWHFIPGLALSAVIT-GVALWGGSIPAVAGAGFSALTLAILLGMVLGNTIYPH
                       10        20         30        40        50        60

489       519       549       579       609       636       666       696
R-QLFNKGIAFTSKYILQTAVVLLGFGLNLMQVMKVGISSLPIIIMTISISLIIAYVL-QKLFKLDKTIATLIGVGSSIC
  :  :|: |  :|:|:  ::| || | : |:  |||| : | ::|:| ::::|  | ||:| |||  : ||| |||||
IWKSCDGGVLFAKQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSSTFLLACFLGQKVFGLDKHTSWLIGAGSSIC
              80        90       100       110       120       130       140

726       756       786       816       840       870       900       930
GGSAIAATAPVINAKDDEVAQAISVIFLFNILAALIFPTLGNFIG--LSDHGFALFAGTAVNDTSSVTATATAWDAINHS
| :|: || ||: |:  :|  |:: : :|  :| :::| :  ::   :|   | :: |: |:: : | |   |  ||
GAAAVLATEPVVKAEASKVTVAVATVVIFGTVAIFLYPAIYPLMSQWFSPETFGIYIGSTVHEVAQVVA---AGHAI-SP
             160       170       180       190       200       210           220

960       990      1020      1050      1077      1107      1134      1164
NTLGGATIVKLTRTLAIIPITIVLSIYHMKQTQKEQSVSVTKI-FPKFVLYFILASLLTTIVASLGF-SLRIFEPLKVLS
:    | | |: | : : |  |:|:   :||     |   :||  | | : ||: :::     |: :    :   |  |
DAENAAVISKMLRVMMLAPFLILLAA-RVKQLSGANSGEKSKITIPWFAILFIVVAIF----NSFHLLPQSVVNMLVTLD
       230       240        250       260       270           280       290

1194      1224      1254      1284      1314      1344      1374      1404
KFFIVMAMGAIGINTNVSKLIKTGGKSILLGAACWLGIIIVSLTMQAILGTW*SCLKLNICNRFHKCYNEDIKRREHYGI
 |:: ||| |:|: |:|| | | | | :|:    :  :|:
TFLLAMAMAALGLTTHVSALKKAGAKPLLMALVLFAWLIVGGGAINYVIQSVIA
            310       320       330       340
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1729

A DNA sequence (GBSx1834) was identified in *S.agalactiae* <SEQ ID 5373> which encodes the amino acid sequence <SEQ ID 5374>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.93    Transmembrane     7 -  23 (   1 -  27)

----- Final Results -----
                 bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5375> which encodes the amino acid sequence <SEQ ID 5376>. Analysis of this protein sequence reveals the following:

```
     Possible site: 40

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.34    Transmembrane    22 -  38 (  13 -  42)

----- Final Results -----
                 bacterial membrane --- Certainty=0.7538(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 56/215 (26%), Positives = 111/215 (51%), Gaps = 5/215 (2%)

Query: 7   VFLTVLVLILIVGAGGLYFWNNHQSLEGKWRTVSLEKQVEKEIEQQLGSQAADMGISAAD 66
           +F+ ++ LIL+    G+ + N+  S+EG WRT S+++++  +  ++L        I  +
Sbjct: 22  LFVFIIFLILLAVLFGVRYRNS--SIEGIWRTTSIDQKLGDDFAKRLTGLHQSPLIDDS- 78

Query: 67  LVKGANMHMNVKNDEAKITVTAQIDEVKFHQAIKTFIDKALEKQLKDQGLTYNDLSEAGK 126
           L+  + M + VKN+   ++ + Q++   F + +  +    L K LK+  L   DLS   +
Sbjct: 79  LLTSSQMILTVKNNNVDLSFSVQVERDIFVKRLAAYHQNELLKTLKENHLVVGDLSSKER 138

Query: 127 KIFDETKITDQQIDQQIDRSFQSAAQAAGGKYNTNTGEMTLPVMDGKVHRLTSVIKV-SH 185
           +I + +     +++  +D++F+  A   GGKYN  TG ++  V+ GKV+R+   I +
Sbjct: 139 QIIENSMPASHELEMILDQAFEKLASQIGGKYNQKTGHLSAVVLKGKVNRILHTIDIKEE 198

Query: 186 INKKANAFYGNIVKNGEKTAYKKEGSKL-ILGNEK 219
           +     +F   ++       Y + G KL +LG+EK
Sbjct: 199 VAAGHTSFSKGLLTPNGYFDYTRFGKKLELLGDEK 233
```

SEQ ID 5374 (GBS288) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 3; MW 53.7kDa).

GBS288d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 8-10; MW 26kDa) and in Figure 183 (lane 3; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 11; MW 51kDa). Purified GBS288d-GST is shown in lane 8 of Figure 237.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1730

A DNA sequence (GBSx1835) was identified in *S.agalactiae* <SEQ ID 5377> which encodes the amino acid sequence <SEQ ID 5378>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3885(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1731

A DNA sequence (GBSx1836) was identified in *S.agalactiae* <SEQ ID 5379> which encodes the amino acid sequence <SEQ ID 5380>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.37   Transmembrane   67 -  83 (  63 -  89)
   INTEGRAL    Likelihood = -3.72   Transmembrane  139 - 155 ( 137 - 158)
   INTEGRAL    Likelihood = -1.54   Transmembrane  115 - 131 ( 114 - 131)

----- Final Results -----
               bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10905> which encodes amino acid sequence <SEQ ID 10906> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1732

A DNA sequence (GBSx1837) was identified in *S.agalactiae* <SEQ ID 5381> which encodes the amino acid sequence <SEQ ID 5382>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4709(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1733

A DNA sequence (GBSx1838) was identified in *S.agalactiae* <SEQ ID 5383> which encodes the amino acid sequence <SEQ ID 5384>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2191(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98427 GB:M63481 20-kDa protein [Streptococcus sanguinis]
 Identities = 119/163 (73%), Positives = 146/163 (89%)

Query: 1   MTTFLGNPVTFTGKQLQVGDIAKDFLLIATDLSQKSLKDFEGKKKVISVVPSIDTGICSK 60
            MTTFLGNPVTFTGKQLQVGD A DF L ATDLS+K+L DF GKKKV+S++PSIDTG+CS
Sbjct: 1   MTTFLGNPVTFTGKQLQVGDTAHDFSLTATDLSKKTLADFAGKKKVLSIIPSIDTGVCST 60

Query: 61  QTRTFNEELSELDNTVVITVSMDLPFAQKRWCSAEGLDNVILLSDFYDHSFGQEYALLMN 120
            QTR FN+ELS+LDNTVVITVS+DLPFAQ +WC+AEG++N ++LSD++DHSFG++YA+L+N
Sbjct: 61  QTRRFNQELSDLDNTVVITVSVDLPFAQGKWCAAEGIENAVMLSDYFDHSFGRDYAVLIN 120

Query: 121 EWHLLTRAVLILDEHNKVTYTEYVDNVNSDVDYEAAINAAKIL 163
            EWHLL RAVL+LDE+N VTY EYVDN+N++ DY+AAI A K L
Sbjct: 121 EWHLLARAVLVLDENNTVTYAEYVDNINTEPDYDAAIAAVKSL 163
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1734

A DNA sequence (GBSx1839) was identified in *S.agalactiae* <SEQ ID 5385> which encodes the amino acid sequence <SEQ ID 5386>. This protein is predicted to be DNA alkylation repair enzyme. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4729(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB40581 GB:AJ010128 DNA alkylation repair enzyme [Bacillus
           cereus]
 Identities = 67/217 (30%), Positives = 119/217 (53%), Gaps = 5/217 (2%)
```

```
Query: 6   SLERKFKAASDKEVSKQQEAYLRHHFKCYGIKSPERRMLYKELIKAAKRQAKIDWQLLDK 65
           +L+  F A  + E ++    Y+++HF   GI++PERR L K++I+      + D+Q++ +
Sbjct: 7   ALQEHFIANQNPEKAEPMARYMKNHFPFLGIQTPERRQLLKDVIQIHTLPDQKDFQVIVR 66

Query: 66  -CWQSDYREYHHFVLDYLLAMSQFLTYNDCSRLEFYARHQQWWDSIDVLTKIF-GNLSLK 123
             W    RE+    LD +      +       LE     + WWD++D +   F GN+ L+
Sbjct: 67  ELWDLPEREFQAAALDMMQKYKMHINETHIPFLEELIVTKSWWDTVDSIVPTFLGNIFLQ 126

Query: 124 DDKVMNL-LSEWSLDQDFWMRRLAIEHQLGFKEKTNTDILSLFILRNTGSQEFFINKAIG 182
             ++++  + +W    + W++R AI  QL +K+K + ++L   I +   S+EFFI KAIG
Sbjct: 127 HPELISAYIPKWIASDNIWLQRAAILFQLKYKQKMDEELLFWVIGQLHSSKEFFIQKAIG 186

Query: 183 WALRDYSKYNKVWVKDFISNHCDELSTLSIREGSKYL 219
           W LR+Y+K     V +++ N  +EL+ LS RE  K++
Sbjct: 187 WVLREYAKTKSDVVWEYVQN--NELAPLSRREAIKHI 221
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1735

A DNA sequence (GBSx1841) was identified in *S.agalactiae* <SEQ ID 5387> which encodes the amino acid sequence <SEQ ID 5388>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA81648 GB:Z27121 unknown [Mycoplasma hominis]
 Identities = 67/281 (23%), Positives = 113/281 (39%), Gaps = 52/281 (18%)

Query: 3   FVFDIDGTLCFDGMS--LSKEIQGILERAQIDYGHRVTFATARSYRDTIGILGDKLSLSK 60
           F  D+DGTL  D  +  +  + +  +++A +  GH V+  T R +R T+ +  +KL L+
Sbjct: 14  FAIDLDGTLLADSANGTVHPKTEEAIKKA-VAQGHIVSIITGRPWRSTLPVY-EKLGLNA 71

Query: 61  IIG-LNGATLHENGHLVDSYYLQSDFFSTIISYCHRHQIPYFVD------EVFNYATYQA 113
           I+G  NGA +H              FF   I+Y   +++ Y +       E+ NYA
Sbjct: 72  IVGNYNGAHIHNPA---------DPFFIPAITYLDLNEVLYILGDEKVKKEITNYAIEGP 122

Query: 114 SKIPFIAYVDPQ-----------KRGELLEVSKIE----------KPIKMVLYFGDQLGR 152
             +  + + DP            K  E + + KI           KP   VL     L R
Sbjct: 123 DWVQLM-HRDPNLERVFGFNQATKFRECINLEKIPLKPTGIVFDVKPDTDVLELLTYLKR 181

Query: 153 ADQMLAELNRFGLSSHFFHEFEKCLYINPIAVDKGKATKKLFG------NRFIAFGNDKN 206
               L E + +         F+    I  I +DKGK    +        +  +A G+  N
Sbjct: 182 RYGDLGEFSSWSKGEGLSPVFD----ITSIGIDKGKVISLIMRYYNIDIDDTVAMGDSYN 237

Query: 207 DISMFDAAHYSVQVGDFDELTPYANLRVSRESVHEGITTLF 247
           D+SM++ A+  V   + + L    +  V +++  EG    F
Sbjct: 238 DLSMYNVANVCVSPANAEPLIKKMSTVVMKQTNKEGAVGYF 278
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1736

A DNA sequence (GBSx1842) was identified in *S.agalactiae* <SEQ ID 5389> which encodes the amino acid sequence <SEQ ID 5390>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2383(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB90005 GB:AE001018 A. fulgidus predicted coding region AF1244
           [Archaeoglobus fulgidus]
 Identities = 22/48 (45%), Positives = 35/48 (72%)

Query: 150 GKSIGELNVWHQTGATIVAIEHEGKFIVSPGPFSVIEQGDHIFFVGDE 197
            GKSIGEL +   +TGAT++A+   + K I+SP P +V+E GD +   +G++
Sbjct: 102 GKSIGELGIRSKTGATVIAVLKKEKTIISPSPETVLEPGDKVVVIGEK 149
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5391> which encodes the amino acid sequence <SEQ ID 5392>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 163/213 (76%), Positives = 196/213 (91%)

Query: 1   MVSEQSEIVTSKYQKIAVAVAQRIANGDYEVGEKLKSRTTIASTFNVSPETARKGLNILA 60
           ++S + EI +SKYQKIA++VAQRIANG+YEVGEKLKSRTTIASTFNVSPETARKGLNILA
Sbjct: 1   VISPKKEITSSKYQKIAISVAQRIANGEYEVGEKLKSRTTIASTFNVSPETARKGLNILA 60

Query: 61  DLQILTLKHGSGAIILSKEKAIEFLNQYETSHSVAILKGKIRDNIKAQQQEMEELATLVD 120
           DL+ILTLKHGSGAI+LSKE+AIEF+NQYE++HS+A+LK KIR+ I  Q + ME++A LV+
Sbjct: 61  DLKILTLKHGSGAIVLSKERAIEFINQYESTHSIAVLKEKIRETINDQGKAMEKMAVLVN 120

Query: 121 DFLLQTRAVSKQYPLAPYEIIVSEDSEHLGKSIGELNVWHQTGATIVAIEHEGKFIVSPG 180
           DFL+Q+++VSKQYPLAPYEII ++DSEH GKSIG LN+WHQTGATIVAIEH G+FIVSPG
Sbjct: 121 DFLMQSQSVSKQYPLAPYEIICNQDSEHFGKSIGVLNIWHQTGATIVAIEHAGQFIVSPG 180

Query: 181 PFSVIEQGDHIFFVGDEDVYARMKTYFNLRMGL 213
           P+SVIE+GDHI+FVGDE V +RMKT+FNLR GL
Sbjct: 181 PYSVIEKGDHIYFVGDESVISRMKTFFNLRKGL 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1737

A DNA sequence (GBSx1844) was identified in *S.agalactiae* <SEQ ID 5393> which encodes the amino acid sequence <SEQ ID 5394>. This protein is predicted to be gls24. Analysis of this protein sequence reveals the following:

```
Possible site: 16
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2855(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9361> which encodes amino acid sequence <SEQ ID 9362> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA86383 GB:U23376 putative 20-kDa protein [Lactococcus lactis]
 Identities = 63/124 (50%), Positives = 84/124 (66%)

Query: 1   MSGGFFSNLKNSVVNSDSVTDGVNVEVGTKEVAVDLDIVVEYGKDIPAIVESIKAIVSQN 60
           + GGFFSNL   ++N+D VT GV+VEVG  +VAVDL +V EY K++P I E IK ++ +
Sbjct: 55  VEGGFFSNLTGKLINTDDVTTGVDVEVGKTQVAVDLKVVTEYRKNVPDIYEKIKEVIRKE 114

Query: 61  VEVMTHLKVVELNANVVDIKTKAEHEADSVTVQDRVSDAAQATGNFASEQAGKAKAAISS 120
           V  MT L+VVE+N  V DIKTK + + D V++QDRV+ AAQ TG F SEQ  K K  +
Sbjct: 115 VAAMTELEVVEVNVTVTDIKTKEQQKEDDVSIQDRVTSAAQTTGKFTSEQVDKVKDKVED 174

Query: 121 GAEK 124
              +K
Sbjct: 175 NTDK 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5395> which encodes the amino acid sequence <SEQ ID 5396>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2534(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/137 (68%), Positives = 108/137 (78%), Gaps = 8/137 (5%)

Query: 1   MSGGFFSNLKNSVVNSDSVTDGVNVEVGTKEVAVDLDIVVEYGKDIPAIVESIKAIVSQN 60
           ++GGFFSN+KN++VNS+SVTDGV+VEVG+KEVAVDL I+VEYGKDIPAI ESIKAIVSQN
Sbjct: 35  VTGGFFSNIKNNLVNSESVTDGVSVEVGSKEVAVDLAIIVEYGKDIPAIAESIKAIVSQN 94

Query: 61  VEVMTHLKVVELNANVVDIKTKAEHEADSVTVQDRVSDAAQATGNFASEQAGKAKAAISS 120
           V+ MTHLKVVE+N NVVDI+TK EHEA SVTVQDRV+ AA +T  F SEQ  K K  IS
Sbjct: 95  VDSMTHLKVVEVNVNVVDIRTKEEHEAASVTVQDRVTSAASSTSQFVSEQTEKLKDTISD 154

Query: 121 GAEKTKEAVSNGTEAAK 137
                     N  EAAK
Sbjct: 155 --------TVNSDEAAK 163
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1738

A DNA sequence (GBSx1845) was identified in *S.agalactiae* <SEQ ID 5397> which encodes the amino acid sequence <SEQ ID 5398>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3393(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1739

A DNA sequence (GBSx1846) was identified in *S.agalactiae* <SEQ ID 5399> which encodes the amino acid sequence <SEQ ID 5400>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3168(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1740

A DNA sequence (GBSx1847) was identified in *S.agalactiae* <SEQ ID 5401> which encodes the amino acid sequence <SEQ ID 5402>. This protein is predicted to be gls24. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA86383 GB:U23376 putative 20-kDa protein [Lactococcus lactis]
 Identities = 95/157 (60%), Positives = 120/157 (75%)

Query: 18  VRGELTFEDKVIEKIVGIAIEHVDGLLAVNGGFFSNLKNSVVNSDSVTDGVNVEVGKKQV 77
            ++G LT+EDKV++KIVG+A+E VDGLL+V GGFFSNL   ++N+D VT GV+VEVGK QV
Sbjct: 27  IKGALTYEDKVVQKIVGLALESVDGLLSVEGGFFSNLTGKLINTDDVTTGVDVEVGKTQV 86

Query: 78  AVDLDIVAEYQKHVPTIFADIKKVVEAEVKRMTDLEVVEVNVNVVDIKTRAQHEEDSVTL 137
            AVDL +V EY+K+VP I+  IK+V+  EV  MT+LEVVEVNV V DIKT+ Q +ED V++
Sbjct: 87  AVDLKVVTEYRKNVPDIYEKIKEVIRKEVAAMTELEVVEVNVTVTDIKTKEQQKEDDVSI 146

Query: 138 QDRVTSAAQATGEFASNQVSNVKSAVGSGVDKVEDMK 174
            QDRVTSAAQ TG+F S QV  VK  V    DK   +K
```

```
Sbjct: 147 QDRVTSAAQTTGKFTSEQVDKVKDKVEDNTDKEARVK 183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5403> which encodes the amino acid sequence <SEQ ID 5404>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3896(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 123/180 (68%), Positives = 158/180 (87%), Gaps = 1/180 (0%)

Query: 1   MTETYIKNTTNNSGTTAVRGELTFEDKVIEKIVGIAIEHVDGLLAVNGGFFSNLKNSVVN 60
           MTETYIKNT+ +  T+A+RG+LT++DKVIEKIVG+A+E+VDGLL VNGGFF+NLK+ +VN
Sbjct: 1   MTETYIKNTSKDL-TSAIRGQLTYDDKVIEKIVGLALENVDGLLGVNGGFFANLKDKLVN 59

Query: 61  SDSVTDGVNVEVGKKQVAVDLDIVAEYQKHVPTIFADIKKVVEAEVKRMTDLEVVEVNVN 120
           ++SV DGVNVEVGKKQVAVDLDIVAEYQKHVPTI+  IK +VE EVKRMTDL+V+EVNV
Sbjct: 60  TESVRDGVNVEVGKKQVAVDLDIVAEYQKHVPTIYDSIKSIVEEEVKRMTDLDVIEVNVK 119

Query: 121 VVDIKTRAQHEEDSVTLQDRVTSAAQATGEFASNQVSNVKSAVGSGVDKVEDMKSEPRVQ 180
           VVDIKT+ Q E + V+LQD+V+  A++T EF S+QV NVK++V +GV+K++D K+EPRV+
Sbjct: 120 VVDIKTKEQFEAEKVSLQDKVSDMARSTSEFTSHQVENVKASVDNGVEKLQDQKAEPRVK 179
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1741

A DNA sequence (GBSx1848) was identified in *S.agalactiae* <SEQ ID 5405> which encodes the amino acid sequence <SEQ ID 5406>. This protein is predicted to be a 6-kDa protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.29   Transmembrane   25 -  41 (  23 -  52)

----- Final Results -----
               bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA86382 GB:U23376 putative 6-kDa protein [Lactococcus lactis]
 Identities = 27/61 (44%), Positives = 45/61 (73%)

Query: 3  EFVRKYRYPLGGAVIGLVLAAMIVTIGFFKTILALVIIVLGAYAGLYVQRTGMLDQFFNK 62
          ++  K RYP+ G ++G ++A  I TIGF+K IL L +I LG Y GL+++++G++DQF N+
Sbjct: 2  DYFEKNRYPIIGGIVGALIAVCIFTIGFWKMILVLFLIGLGIYIGLFLKKSGIIDQFINR 61

Query: 63 R 63
          +
Sbjct: 62 K 62
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5407> which encodes the amino acid sequence <SEQ ID 5408>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.73   Transmembrane   11 -  27 (   6 -  50)
   INTEGRAL    Likelihood = -7.11   Transmembrane   33 -  49 (  27 -  50)

----- Final Results -----
               bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 28/61 (45%), Positives = 48/61 (77%)

Query: 3  EFVRKYRYPLGGAVIGLVLAAMIVTIGFFKTILALVIIVLGAYAGLYVQRTGMLDQFFNKR 63
           EF  K++YP+ G ++GL++A +++  G FKT+LA++ I+LG Y GLY ++TG++DQF N++
Sbjct: 2  EFYEKFKYPIIGGLVGLIIAILLMAFGLFKTLLAIIFIILGIYGGLYAKKTGIIDQFLNRK 62
```

A related GBS gene <SEQ ID 8891> and protein <SEQ ID 8892> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     12.56
GvH: Signal Score (-7.5): -1.11
     Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -9.29 threshold:  0.0
   INTEGRAL    Likelihood = -9.29   Transmembrane   25 -  41 (  23 -  52)
   PERIPHERAL  Likelihood = 12.25      44
 modified ALOM score:   2.36

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
44.3/73.8% over 60aa
  Lactococcus lactis
  EGAD|42618| putative 6-kDa protein Insert characterized
  GP|727435|gb|AAA86382.1||U23376 putative 6-kDa protein Insert characterized

ORF01006(307 - 489 of 792)
EGAD|42618|45008(2   -   62   of   62)   putative   6-kDa   protein   {Lactococcus
lactis}GP|727435|gb|AAA86382.1||U23376 putative 6-kDa protein {La
ctococcus lactis}
%Match = 11.6
%Identity = 44.3  %Similarity = 73.8
Matches = 27  Mismatches = 16  Conservative Sub.s = 18

159       189       219       249       279       309       339       369
TNVPEQLEHIQSDVELGLKEFFGLEKKMNTRVFVKQVEEENVGNAKTNKSRVE*ESNMSEFVRKYRYPLGGAVIGLVLAA
                                                           ::  | |||: | ::| ::|
                                                          MDYFEKNRYPIIGGIVGALIAV
                                                                  10        20

399       429       459       489       519       549       579       609
MIVTIGFFKTILALVIIVLGAYAGLYVQRTGMLDQFFNKRK*NFSFIFILHYLNKRKRNYYD*NLHQKHN*QFWHDSCSW
 | ||||:| || | :| || | ||:::::|::||| |::
CIFTIGFWKMILVLFLIGLGIYIGLFLKKSGIIDQFINRK
                40        50        60
```

SEQ ID 5406 (GBS14) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 4; MW 33.3kDa). The GBS14-GST fusion product was purified (Figure 190, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 263), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1742

A DNA sequence (GBSx1849) was identified in *S.agalactiae* <SEQ ID 5409> which encodes the amino acid sequence <SEQ ID 5410>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-18.63    Transmembrane    61 -  77 (  51 -  83)
   INTEGRAL    Likelihood = -7.91    Transmembrane    10 -  26 (   7 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.8451(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5411> which encodes the amino acid sequence <SEQ ID 5412>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-16.19    Transmembrane    71 -  87 (  63 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.7474(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/193 (45%), Positives = 127/193 (65%), Gaps = 4/193 (2%)

Query: 1   MSKGLKSLYTLLGLISLTLLGFVAVISKQHIYLP-SFNWLDWDFN-LPSPIDVGMYHYFF 58
           MSK LK  Y L+GL+ L++ G+V  I+  +IYLP S+ WL W  +  P+ +D  + +Y+F
Sbjct: 9   MSKLLKISYCLVGLVLLSVFGWVVGITGGYIYLPYSYRWLSWGMDSFPNLLDSALSYYYF 68

Query: 59  WGALVLFVIVLLAILVVLFYPRRYTEYKLA--DKTGKLMLKKSAIEGFVKTEVLKTGLMK 116
           W ALVLFVI  LA+LV++ YPR YTE +L   +K G L+LKKSAIE +V T +   GLM
Sbjct: 69  WTALVLFVITFLALLVIILYPRIYTEVQLRHKNKKGTLLLKKSAIESYVATAIQTAGLMP 128

Query: 117 SPSVTAHLYKKKVKVDVKGLLTSRTNVPEQLEHIQSDVELGLKEFFGLEKKMNTRVFVKQ 176
           +P+VTA LYK+K  + VKG L SR  V +Q+  ++  +E GL EFFG+   +N +V+VK
Sbjct: 129 NPTVTAKLYKRKFNIIVKGRLASRVAVADQISGVKEGIEKGLTEFFGINYPVNFKVYVKD 188

Query: 177 VEEENVGNAKTNK 189
           + + +  +   N+
Sbjct: 189 IADSDRKHITRNR 201
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1743

A DNA sequence (GBSx1850) was identified in *S.agalactiae* <SEQ ID 5413> which encodes the amino acid sequence <SEQ ID 5414>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.82    Transmembrane   56 -  72 (  52 -  81)
   INTEGRAL    Likelihood = -6.42    Transmembrane    4 -  20 (   1 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12244  GB:Z99106  similar  to  hypothetical  proteins  from  B.  subtilis  [Bacillus
subtilis]
 Identities = 31/76 (40%), Positives = 48/76 (62%)

Query: 1  MSLIWSLIVGAIIGAIAGAVTNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLAGMALI 60
          +S + SL+V  +IG I  A+       G   +++AGL+G+++G  LLGTWGP LAG A+
Sbjct: 2  LSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAIF 61

Query: 61 PSIVGAIIVVIVTSFV 76
          P+I+GA I V +   +
Sbjct: 62 PAIIGAAIFVFLLGLI 77
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5415> which encodes the amino acid sequence <SEQ ID 5416>. Analysis of this protein sequence reveals the following:

```
     Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.59    Transmembrane   60 -  76 (  56 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12244  GB:Z99106  similar  to  hypothetical  proteins  from  B.  subtilis  [Bacillus
subtilis]
 Identities = 28/76 (36%), Positives = 47/76 (61%)

Query: 1  MGLIWTLIVGALIGVIAGALTKKGGSMGWIANIAAGLVGSSVGQALLGSWGPSLAGMSLI 60
          +  + +L+V  +IG+I  A+       G   ++ AGL+G+ +G  LLG+WGPSLAG ++
Sbjct: 2  LSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAIF 61

Query: 61 PSVIGAVIVVMITSFV 76
          P++IGA I V +   +
Sbjct: 62 PAIIGAAIFVFLLGLI 77
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 63/82 (76%), Positives = 74/82 (89%)

Query: 1  MSLIWSLIVGAIIGAIAGAVTNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLAGMALI 60
          M LIW+LIVGA+IG IAGA+T KGGSMGWIANI AGLVGS VGQ+LLG+WGP LAGM+LI
Sbjct: 1  MGLIWTLIVGALIGVIAGALTKKGGSMGWIANIAAGLVGSSVGQALLGSWGPSLAGMSLI 60

Query: 61 PSIVGAIIVVIVTSFVLGKMNN 82
          PS++GA+IVV++TSFVL K NN
Sbjct: 61 PSVIGAVIVVMITSFVLNKTNN 82
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1744

A DNA sequence (GBSx1851) was identified in *S.agalactiae* <SEQ ID 5417> which encodes the amino acid sequence <SEQ ID 5418>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane   88 - 104 (  84 - 111)
   INTEGRAL    Likelihood = -8.07   Transmembrane   29 -  45 (  27 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12244  GB:Z99106  similar  to  hypothetical  proteins  from  B.  subtilis  [Bacillus
subtilis]
 Identities = 29/77 (37%), Positives = 47/77 (60%)

Query: 31    IMGLIWSLIVGAIIGAIAGAITNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLADMAL 90
             ++  + SL+V  +IG I  AI       G   +++AGL+G+++G  LLGTWGP LA  A+
Sbjct: 1     MLSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAI 60

Query: 91    IPSIVGAIIVIIVTSFV 107
              P+I+GA I + +   +
Sbjct: 61    FPAIIGAAIFVFLLGLI 77
```

There is also homology to SEQ ID 5416:

```
Identities = 60/79 (75%), Positives = 72/79 (90%)

Query: 32    MGLIWSLIVGAIIGAIAGAITNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLADMALI 91
             MGLIW+LIVGA+IG IAGA+T KGGSMGWIANI AGLVGS VGQ+LLG+WGP LA M+LI
Sbjct: 1     MGLIWTLIVGALIGVIAGALTKKGGSMGWIANIAAGLVGSSVGQALLGSWGPSLAGMSLI 60

Query: 92    PSIVGAIIVIIVTSFVLGK 110
             PS++GA+IV+++TSFVL K
Sbjct: 61    PSVIGAVIVVMITSFVLNK 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1745

A DNA sequence (GBSx1852) was identified in *S.agalactiae* <SEQ ID 5419> which encodes the amino acid sequence <SEQ ID 5420>. This protein is predicted to be ATP-dependent DNA helicase Rep (uvrD). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1364(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9863> which encodes amino acid sequence <SEQ ID 9864> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD51119 GB:AF176554 DNA helicase PcrA [Leuconostoc citreum]
 Identities = 414/764 (54%), Positives = 537/764 (70%), Gaps = 23/764 (3%)

Query: 6   VEMNPLIIGMNDKQAEAVQTTDGPLLIMAGAGSGKTRVLTHRIAYLIDEKYVNPWNILAI 65
           + +  L  GMN+KQAEAVQTT+GPLLIMAGAGSGKTRVLTHRIA+L+ +  V PW ILAI
Sbjct: 1   MSVETLTNGMNNKQAEAVQTTEGPLLIMAGAGSGKTRVLTHRIAHLVQDLNVFPWRILAI 60

Query: 66  TFTNKAAREMRERAIAL--NPATQDTLIATFHSMCVRILRREADYIGYNRNFTIVDPGEQ 123
           TFTNKAAREMRER  AL      +D  ++TFH++ VRILRR+ + IG  +NFTI+D   Q
Sbjct: 61  TFTNKAAREMRERIAALLSEDVARDIWVSTFHALAVRILRRDGEAIGLAKNFTIIDTSAQ 120

Query: 124 RTLMKRIIKQLNLDTKKWNERSILGTISNAKNDLLDEIAYEKQAGDMYTQVIAKCYKAYQ 183
           RTLMKR+I  LNLDT +++ R+ILG ISNAKND+L    Y K A + + + +A+ Y AYQ
Sbjct: 121 RTLMKRVINDLNLDTNQYDPRTILGMISNAKNDMLQPRDYAKAADNAFQETVAEVYTAYQ 180

Query: 184 EELRRSEAMDFDDLIMMTLRLFDQNKDVLAYYQQRYQYIHVDEYQDTNHAQYQLVKLLAS 243
            EL+RS+++DFDDLIM+T+ LF    DVLA YQQ+++Y+HVDEYQDTN AQY +V LLA
Sbjct: 181 AELKRSQSVDFDDLIMLTIDLFQSAPDVLARYQQQFEYLHVDEYQDTNDAQYTIVNLLAQ 240

Query: 244 RFKNICVVGDADQSIYGWRGADMQNILDFEKDYPQAKVVLLEENYRSTKKILQAANNVIN 303
           R KN+ VVGDADQSIYGWRGA+M NIL+FEKDYP A  V+LE+NYRST+ IL AAN VIN
Sbjct: 241 RSKNLAVVGDADQSIYGWRGANMNNILNFEKDYPNAHTVMLEQNYRSTQNILDAANAVIN 300

Query: 304 HNKNRRPKKLWTQNDEGEQIVYHRANNEQEEAVFVASTIDNIVREQGKNFKDFAVLYRTN 363
           HN  R PKKLWT+N +G+QI Y+RA  E +EA F+ S I  +   +   + DFAVLYRTN
Sbjct: 301 HNNERVPKKLWTENGKGDQITYYRAQTEHDEANFILSNIQQLRETKHMAYSDFAVLYRTN 360

Query: 364 AQSRTIEEALLKSNIPYTMVGGTKFYSRKEIRDVIAYLNILANTSDNISFERIVNEPKRG 423
           AQSR IEE+L+K+N+PY+MVGG KFY RKEI D++AY++++ N  DN +FER+VNEPKRG
Sbjct: 361 AQSRNIEESLVKANMPYSMVGGHKFYERKEILDIMAYMSLITNPDDNAAFERVVNEPKRG 420

Query: 424 VGPGTLEKIRSFAYEQSMSLLDASSNVMMSP-LKGKAAQAVWDLANLILTLRSNLDSLTV 482
           +G  +L ++R  A   ++S + A  ++ ++P +  KAA      A ++  LR   + L V
Sbjct: 421 LGATSLTRLRELANRLNVSYMKAIGSIELAPSITTKAASKFLTFAEMMHNLRQQSEFLNV 480

Query: 483 TEITENLLDKTGYLEALQVQNTLESQARIENIEEFLSVTKNFDDNPEITVEGETGLDRLS 542
           TE+TE ++ ++GY + L  +N  +SQAR+EN+EEFLSVTK FDD  +   E    +D ++
Sbjct: 481 TELTELVMTQSGYRQMLAEKNDPDSQARLENLEEFLSVTKEFDD--KYQPEDPESIDPVT 538

Query: 543 RFLNDLALIADTDDSATETAEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRAIEDADEL 602
            FL   AL++D DD       VTLMTLHAAKGLEFPVVFLIG++EG+FPLSRA+ D D L
Sbjct: 539 DFLGTTALMSDLDDFEEGDGAVTLMTLHAAKGLEFPVVFLIGLKEGIFPLSRAMMDEDLL 598

Query: 603 EEERRLAYVGITRAEQILFLTNANTRTLFGKTSYNRPTRFIREIDDELIQ--YQGLARPV 660
           EEERRLAYVGITRA + LFLTNA +R L+G+T  N P+RFI EI  EL++  Y GL+R
Sbjct: 599 EEERRLAYVGITRAMKKLFLTNAFSRLLYGRTQANEPSRFIAEISPELLETAYSGLSRDK 658

Query: 661 NSSFGVKYSKEQPTQFGQGMSLQQALQARKSNSQSQVTAQLQALN-ANNSHETSWEIGDV 719
                + + ++                 R + +  Q T   +  N      +TSW  GD
Sbjct: 659 TQKKTLPFDRK---------------MQRATATTYQATPVTKITNGVTGGDQTSWSTGDK 703

Query: 720 ATHKKWGDGTVLEVSGSGKTQELKINFPGIGLKKLLASVAPISK 763
            +HKKWG GTV+ VSG    QELK+ FP  G+K+LLA+ API K
Sbjct: 704 VSHKKWGVGTVISVSGRADDQELKVAFPSEGVKQLLAAFAPIQK 747
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5421> which encodes the amino acid sequence <SEQ ID 5422>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0214(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 622/772 (80%), Positives = 699/772 (89%), Gaps = 15/772 (1%)

Query: 8   MNPLIIGMNDKQAEAVQTTDGPLLIMAGAGSGKTRVLTHRIAYLIDEKYVNPWNILAITF 67
           MNPL+ GMND+QA+AVQTT+GPLLIMAGAGSGKTRVLTHRIAYLIDEK+VNPWNILAITF
Sbjct: 1   MNPLLNGMNDRQAQAVQTTEGPLLIMAGAGSGKTRVLTHRIAYLIDEKFVNPWNILAITF 60

Query: 68  TNKAAREMRERAIALNPATQDTLIATFHSMCVRILRREADYIGYNRNFTIVDPGEQRTLM 127
           TNKAAREM+ERA+ALNPAT+DTLIATFHSMCVRILRREAD+IGYNRNFTIVDPGEQRTLM
Sbjct: 61  TNKAAREMKERALALNPATKDTLIATFHSMCVRILRREADHIGYNRNFTIVDPGEQRTLM 120

Query: 128 KRIIKQLNLDTKKWNERSILGTISNAKNDLLDEIAYEKQAGDMYTQVIAKCYKAYQEELR 187
           KRI+KQLN+D KKWNERSILGTISNAKNDLLDE  YE QA DMY+Q++A+CYKAYQEELR
Sbjct: 121 KRILKQLNIDPKKWNERSILGTISNAKNDLLDEKGYEAQAADMYSQIVARCYKAYQEELR 180

Query: 188 RSEAMDFDDLIMMTLRLFDQNKDVLAYYQQRYQYIHVDEYQDTNHAQYQLVKLLASRFKN 247
           RSEA+DFDDLIMMTLRLFD N DVLAYYQQRYQYIHVDEYQDTNHAQYQL+KLLASRFKN
Sbjct: 181 RSEALDFDDLIMMTLRLFDANPDVLAYYQQRYQYIHVDEYQDTNHAQYQLIKLLASRFKN 240

Query: 248 ICVVGDADQSIYGWRGADMQNILDFEKDYPQAKVVLLEENYRSTKKILQAANNVINHNKN 307
           ICVVGDADQSIYGWRGADMQNILDFEKDYP AKVVLLEENYRSTKKILQAAN+VIN+N+N
Sbjct: 241 ICVVGDADQSIYGWRGADMQNILDFEKDYPDAKVVLLEENYRSTKKILQAANDVINNNRN 300

Query: 308 RRPKKLWTQNDEGEQIVYHRANNEQEEAVFVASTIDNIVREQGKNFKDFAVLYRTNAQSR 367
           RRPKKLWTQN +GEQ+VY+RAN+E++EAVFVASTI N+ +E GKNFKDFAVLYRTNAQSR
Sbjct: 301 RRPKKLWTQNADGEQLVYYRANDERDEAVFVASTISNMSQELGKNFKDFAVLYRTNAQSR 360

Query: 368 TIEEALLKSNIPYTMVGGTKFYSRKEIRDVIAYLNILANTSDNISFERIVNEPKRGVGPG 427
           TIEEALLKSNIPYTMVGGTKFYSRKEIRD+IAYL I+AN +DNISFERIVNEPKRGVGPG
Sbjct: 361 TIEEALLKSNIPYTMVGGTKFYSRKEIRDLIAYLTIVANPADNISFERIVNEPKRGVGPG 420

Query: 428 TLEKIRSFAYEQSMSLLDASSNVMMSPLKGKAAQAVWDLANLILTLRSNLDSLTVTEITE 487
           TL+K+R FAYE   SLL+A+SN++MSPLKGKAAQA+ DLAN++  LR +LD +++T++ E
Sbjct: 421 TLDKLRQFAYESDQSLLEAASNLLMSPLKGKAAQAIMDLANILGQLRQDLDQMSITDLAE 480

Query: 488 NLLDKTGYLEALQVQNTLESQARIENIEEFLSVTKNFDDNPEITVEGETGLDRLSRFLND 547
            LL+KTGYL++L++QNTLESQARIENIEEFLSVTKNFD++     E ETG+DRL RFLND
Sbjct: 481 ALLEKTGYLDSLRLQNTLESQARIENIEEFLSVTKNFDESSASQEEDETGVDRLGRFLND 540

Query: 548 LALIADTDDSATETAEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRAIEDADELEEERR 607
           LALIADTDDS  E AEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRA ED DELEEERR
Sbjct: 541 LALIADTDDSQAEAAEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRASEDPDELEEERR 600

Query: 608 LAYVGITRAEQILFLTNANTRTLFGKTSYNRPTRFIREIDDELIQYQGLARPVNSSFGVK 667
           LAYVGITRAE++LF+TNANTRTLFGK+SYNRPTRF++EI +EL+ Y+GLARP  SSFGV+
Sbjct: 601 LAYVGITRAEEVLFMTNANTRTLFGKSSYNRPTRFLKEISEELLSYKGLARPAQSSFGVR 660

Query: 668 YSKEQPTQFGQGMSLQQALQARKSNSQSQVTAQ-LQA-------------LNANNS-HET 712
           +S E  TQFGQGMSL +ALQARK+ +Q + +AQ +QA               +N+S  E
Sbjct: 661 FSTETHTQFGQGMSLSEALQARKAQAQVRQSAQPMQAHTIPSASTSSVLPFGSNSSVEEV 720

Query: 713 SWEIGDVATHKKWGDGTVLEVSGSGKTQELKINFPGIGLKKLLASVAPISKK 764
           +W+IGD+A HKKWGDGTVLEVSGSGKT ELKI FP +GLKKLLASVAPI KK
Sbjct: 721 TWQIGDIAHHKKWGDGTVLEVSGSGKTMELKIKFPEVGLKKLLASVAPIEKK 772
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1746

A DNA sequence (GBSx1853) was identified in *S.agalactiae* <SEQ ID 5423> which encodes the amino acid sequence <SEQ ID 5424>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4741(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA88579 GB:M14339 unknown [Streptococcus pneumoniae]
 Identities = 43/57 (75%), Positives = 50/57 (87%)

Query: 41 AHGGYLFTLCDQVSGLVAISTGYEAVTLQSNINYLRAGRLDDLLTVIGTCVHNGRTT 97
           AHGGYLFTLCDQ+SGLV IS G + VTLQS+INYL+AG+LDD+LT+ G CVH GRTT
Sbjct: 1  AHGGYLFTLCDQISGLVVISLGLDGVTLQSSINYLKAGKLDDVLTIKGECVHQGRTT 57
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5425> which encodes the amino acid sequence <SEQ ID 5426>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 57/97 (58%), Positives = 74/97 (75%)

Query: 2   KFNLEQVKVFENYEIENWEEGQVTLTTKVVDSSLNYYGNAHGGYLFTLCDQVSGLVAIST 61
            +  L  + +F+NY+IE  E+G + L+T+V +++LNYYGNAHGGYLFTLCDQV GLVA +T
Sbjct: 7   EMTLNVISIFDNYQIELAEKGHLILSTEVTETALNYYGNAHGGYLFTLCDQVGGLVARTT 66

Query: 62  GYEAVTLQSNINYLRAGRLDDLLTVIGTCVHNGRTTK 98
            G E+VTLQ+N NYL+AG   D L V G  VH GRTT+
Sbjct: 67  GVESVTLQANANYLKAGHKGDKLMVEGRLVHGGRTTQ 103
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1747

A DNA sequence (GBSx1854) was identified in *S.agalactiae* <SEQ ID 5427> which encodes the amino acid sequence <SEQ ID 5428>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3187(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1748

A DNA sequence (GBSx1855) was identified in *S.agalactiae* <SEQ ID 5429> which encodes the amino acid sequence <SEQ ID 5430>. This protein is predicted to be uracil permease (uraA). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65    Transmembrane  122 - 138 ( 117 - 146)
   INTEGRAL    Likelihood = -8.65    Transmembrane  212 - 228 ( 204 - 233)
   INTEGRAL    Likelihood = -7.32    Transmembrane   60 -  76 (  49 -  80)
   INTEGRAL    Likelihood = -6.53    Transmembrane  149 - 165 ( 145 - 172)
   INTEGRAL    Likelihood = -6.48    Transmembrane  402 - 418 ( 401 - 420)
   INTEGRAL    Likelihood = -4.04    Transmembrane  422 - 438 ( 420 - 445)
   INTEGRAL    Likelihood = -3.72    Transmembrane  365 - 381 ( 364 - 385)
   INTEGRAL    Likelihood = -3.40    Transmembrane  184 - 200 ( 182 - 202)
   INTEGRAL    Likelihood = -3.08    Transmembrane  346 - 362 ( 345 - 363)
   INTEGRAL    Likelihood = -1.38    Transmembrane  260 - 276 ( 260 - 276)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9865> which encodes amino acid sequence <SEQ ID 9866> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA53697 GB:X76083 uracil permease [Bacillus caldolyticus]
 Identities = 208/416 (50%), Positives = 291/416 (69%), Gaps = 11/416 (2%)

Query: 32  LLDIDEKPELFQGLLLSFQHVFAMFGATILVPLILGMPVSVALFASGCGTLIYQVATKFK 91
            +LDI ++P + Q + LS QH+FAMFGATILVP ++G+  S+AL  SG GTL + + TK++
Sbjct: 5   VLDIQDRPTVGQWITLSLQHLFAMFGATILVPYLVGLDPSIALLTSGLGTLAFLLITKWQ 64

Query: 92  VPVYLGSSFAYITAMALAMKQMHGDISAAQTGILFVGLIYVVVATVIKFVGNSWVDKILP 151
            VP YLGSSFAYI  +  A  +  G   AA  G    GL+Y VVA +IK  G  WV K+LP
Sbjct: 65  VPAYLGSSFAYIAPIIAA--KTAGGPGAAMIGSFLAGLVYGVVALIIKKAGYRWVMKLLP 122

Query: 152 PIIIGPMIIVIGLGLANSAVTNA--GFVAKGDWRKMLVAVVTFLIAAFINTKGKGFIKII 209
            P+++GP+IIVIGLGLA +AV  A  G   K       VA+VT       +   +G + +I
Sbjct: 123 PVVVGPVIIVIGLGLAGTAVGMAMNGPDGKYSLLHFSVALVTLAATIVCSVLARGMLSLI 182

Query: 210 PFLFAIIGGYILSIILGLVDLSPVEKAAWFELPKFYLPFKTGLFHSYKLYFGPEMLAIL- 268
            P L  I+ GY+ ++ +GLVDLS V  A WFE P F +PF       Y +    E++ ++
Sbjct: 183 PVLVGIVVGYLYALAVGLVDLSKVAAAKWFEWPDFLIPFA-----DYPVRVTWEIVMLMV 237

Query: 269 PISIVTIAENIGDHTVLGQICGRNFLKKPGLNRLLIGDGLATAFSALIGGPAETTYGENT 328
            P++IVT++E+IG   VL ++ GR+ ++KPGL+R ++GDG AT  SAL+GGP +TTYGEN
Sbjct: 238 PVAIVTLSEHIGHQLVLSKVVGRDLIQKPGLHRSILGDGTATMISALLGGPPKTTYGENI 297

Query: 329 GVIGMTRIASVTVIRNAAFIAIAFSFFGKFTALISTIPSAVLGGMAILLYGVIASNGLKV 388
            GV+ +TR+ SV V+  AA IAIAF F GK TALIS+IP+ V+GG++ILL+G+IAS+GL++
Sbjct: 298 GVLAITRVYSVYVLAGAAVIAIAFGFVGKITALISSIPTPVMGGVSILLFGIIASSGLRM 357

Query: 389 LIENRVNFAEVRNLIIASSMLVLGLGGAVLDLG-ALTLSGTALSAIVGIILNLILP 443
            LI++RV+F + RNL+IAS +LV+G+GGAVL +  +  ++G ALSAIVG++LNLILP
Sbjct: 358 LIDSRVDFGQTRNLVIASVILVIGIGGAVLKISDSFQITGMALSAIVGVLLNLILP 413
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5431> which encodes the amino acid sequence <SEQ ID 5432>. Analysis of this protein sequence reveals the following:

```
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood =-10.72   Transmembrane  177 - 193 ( 171 - 206)
     INTEGRAL    Likelihood = -8.55   Transmembrane  313 - 329 ( 304 - 339)
     INTEGRAL    Likelihood = -8.17   Transmembrane  154 - 170 ( 152 - 175)
     INTEGRAL    Likelihood = -7.91   Transmembrane  376 - 392 ( 374 - 395)
     INTEGRAL    Likelihood = -7.48   Transmembrane   25 -  41 (  22 -  43)
     INTEGRAL    Likelihood = -5.84   Transmembrane  120 - 136 ( 116 - 142)
     INTEGRAL    Likelihood = -4.99   Transmembrane   96 - 112 (  90 - 117)
     INTEGRAL    Likelihood = -3.29   Transmembrane  339 - 355 ( 338 - 360)
     INTEGRAL    Likelihood = -1.91   Transmembrane  396 - 412 ( 396 - 413)

----- Final Results -----
                 bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB89870 GB:AJ132624 uracil transporter [Lactococcus lactis]
 Identities = 294/421 (69%), Positives = 359/421 (84%), Gaps = 5/421 (1%)

Query: 3   DVIYDVEEVPKAGMLVGLSFQHLFAMFGATVLVPILVGIDPSVALLSSGLGTLAHLSVTK 62
           D+I  V+E P A    GLSFQHLFAMFG+TVLVPILVGI+P++ALLSSGLGTLAH+SVTK
Sbjct: 5   DIILKVDEKPAASQWFGLSFQHLFAMFGSTVLVPILVGINPAIALLSSGLGTLAHMSVTK 64

Query: 63  FKIPAYMGSSFAYIAAMQLLMKTNGIGAVAQGAMTGGLVYLIVALIVKAIGNDWIDNILP 122
           FK+PAYMGSSFAYI AM LLMK  G+ A+AQGAMTGGLVYLIVALIVK  G  WID +LP
Sbjct: 65  FKVPAYMGSSFAYIGAMTLLMKNGGMPAIAQGAMTGGLVYLIVALIVKFAGKGWIDKVLP 124

Query: 123 PIVVGPIVMVIGLSLASTAVNDVMLKN----GNYNLTYLVIGLVTLLSVIFFNIYGKGIV 178
           PIVVGPIVMVIGLSLA TA+ND M  +      Y+L Y++I L+T+LS++ ++IYGKG +
Sbjct: 125 PIVVGPIVMVIGLSLAPTAINDAMYTDVANLKGYSLAYIIIALITVLSIVVYSIYGKGFL 184

Query: 179 AIVPLLLGLLVGYVVALLVGVLTGQEIVDFTNVAQAKWFSIPSVEIPFLTYGVKFYPSAI 238
           ++VP+LLG++ GYV A+++G +TG  IV FT ++QAKW ++P +EIPF +Y   FYPSAI
Sbjct: 185 SVVPILLGIITGYVAAMIIGKITGMNIVSFTGISQAKWLTLPPMEIPFASYKWAFYPSAI 244

Query: 239 LTMAPIAFVTMTEHFGHIMVLNSLTKRDYFKDPGLEKTLTGDGFAQIIAGFLGAPPVTSY 298
           LTMAPIAFVTMTEHFGHIMVLNSLTK+DYFK+PGLEKTLTGDG AQIIAGF+GAPPVTSY
Sbjct: 245 LTMAPIAFVTMTEHFGHIMVLNSLTKKDYFKEPGLEKTLTGDGLAQIIAGFIGAPPVTSY 304

Query: 299 GENIGVMALNKIFSVYVIAGAAVIAALLSFIGKVSALIQSIPTPVIGGISVALFGVIASS 358
           GENIGVMA+ KI S+YVIAGAAV+A ++SF+GK++AL+QSIP PVIGG S+ALFGVIA+S
Sbjct: 305 GENIGVMAITKIHSIYVIAGAAVLAIVVSFVGKITALLQSIPAPVIGGASIALFGVIAAS 364

Query: 359 GLKILIESKVDMDNKKNLLIASVILVSGIGGLMLQV-NGLQISGVAFSTLLGIILYQVLPE 418
           GLKIL+E+KVD D K+NLLI+SV+LV GIGG+++ +   LQIS VA +T+LGI+L  VLP+
Sbjct: 365 GLKILVENKVDFDIKRNLLISSVVLVIGIGGMIINITQNLQISSVAIATILGIVLNLVLPK 425
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 186/425 (43%), Positives = 282/425 (65%), Gaps = 17/425 (4%)

Query: 30  NLLLDIDEKPELFQGLLLSFQHVFAMFGATILVPLILGMPVSVALFASGCGTLIYQVATK 89
           +++ D++E P+    + LSFQH+FAMFGAT+LVP+++G+  SVAL +SG GTL +   TK
Sbjct: 3   DVIYDVEEVPKAGMLVGLSFQHLFAMFGATVLVPILVGIDPSVALLSSGLGTLAHLSVTK 62

Query: 90  FKVPVYLGSSFAYITAMALAMKQMHGDISAAQTGILFVGLIYVVVATVIKFVGNSWVDKI 149
           FK+P Y+GSSFAYI AM L MK     I A   G +  GL+Y++VA ++K +GN W+D I
Sbjct: 63  FKIPAYMGSSFAYIAAMQLLMKT--NGIGAVAQGAMTGGLVYLIVALIVKAIGNDWIDNI 120

Query: 150 LPPIIIGPMIIVIGLGLANSAVTNAGFVAKGDWRK--MLVAVVTFLIAAFINTKGKGFIK 207
           LPPI++GP+++VIGL LA++AV +   +  G++    +++ +VT L   F N  GKG +
```

```
Sbjct: 121 LPPIVVGPIVMVIGLSLASTAVNDV-MLKNGNYNLTYLVIGLVTLLSVIFFNIYGKGIVA 179

Query: 208 IIPFLFAIIGGYILSIILG------LVDLSPVEKAAWFELPKFYLPFKTGLFHSYKLYFG 261
           I+P L  ++ GY++++++G      +VD + V +A WF +P   +PF T     Y + F
Sbjct: 180 IVPLLLGLLVGYVVALLVGVLTGQEIVDFTNVAQAKWFSIPSVEIPFLT-----YGVKFY 234

Query: 262 PE-MLAILPISIVTIAENIGDHTVLGQICGRNFLKKPGLNRLLIGDGLATAFSALIGGPA 320
           P  +L + PI+ VT+ E+ G   VL  +  R++ K PGL + L GDG A   +  +G P
Sbjct: 235 PSAILTMAPIAFVTMTEHFGHIMVLNSLTKRDYFKDPGLEKTLTGDGFAQIIAGFLGAPP 294

Query: 321 ETTYGENTGVIGMTRIASVTVIRNAAFIAIAFSFFGKFTALISTIPSAVLGGMAILLYGV 380
            T+YGEN GV+ + +I SV VI  AA IA   SF GK +ALI +IP+ V+GG+++ L+GV
Sbjct: 295 VTSYGENIGVMALNKIFSVYVIAGAAVIAALLSFIGKVSALIQSIPTPVIGGISVALFGV 354

Query: 381 IASNGLKVLIENRVNFAEVRNLIIASSMLVLGLGGAVLDLGALTLSGTALSAIVGIILNL 440
           IAS+GLK+LIE++V+    +NL+IAS +LV G+GG +L +  L +SG A S ++GIIL
Sbjct: 355 IASSGLKILIESKVDMDNKKNLLIASVILVSGIGGLMLQVNGLQISGVAFSTLLGIILYQ 414

Query: 441 ILPKE 445
           +LP++
Sbjct: 415 VLPEK 419
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1749

A DNA sequence (GBSx1856) was identified in *S.agalactiae* <SEQ ID 5433> which encodes the amino acid sequence <SEQ ID 5434>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3863(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1750

A DNA sequence (GBSx1857) was identified in *S.agalactiae* <SEQ ID 5435> which encodes the amino acid sequence <SEQ ID 5436>. This protein is predicted to be sodium/alanine symporter. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.88   Transmembrane  191 - 207 ( 184 - 214)
   INTEGRAL    Likelihood = -8.97   Transmembrane  151 - 167 ( 148 - 171)
   INTEGRAL    Likelihood = -8.39   Transmembrane  217 - 233 ( 216 - 238)
   INTEGRAL    Likelihood = -6.74   Transmembrane  312 - 328 ( 310 - 333)
   INTEGRAL    Likelihood = -6.26   Transmembrane  357 - 373 ( 349 - 376)
   INTEGRAL    Likelihood = -5.10   Transmembrane  424 - 440 ( 422 - 441)
   INTEGRAL    Likelihood = -5.04   Transmembrane  396 - 412 ( 390 - 417)
   INTEGRAL    Likelihood = -0.37   Transmembrane   25 -  41 (  25 -  41)

----- Final Results -----
```

```
                   bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9867> which encodes amino acid sequence <SEQ ID 9868> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22541 GB:U32770 amino acid carrier protein, putative
            [Haemophilus influenzae Rd]
 Identities = 255/443 (57%), Positives = 333/443 (74%), Gaps = 4/443 (0%)

Query: 11  TLFTHINSFVWGPPLLALLVGTGIYLSFRLGFIQLRQLSRAFKLIFREDNG-QGDISSYA 69
           ++ + I+SF+WG PLL LL GTG+YL+ RLGFIQ+R L RA   +F++D G +GD+SS+A
Sbjct: 5   SILSAIDSFIWGAPLLILLSGTGLYLTLRLGFIQIRYLPRALGYLFKKDKGGKGDVSSFA 64

Query: 70  ALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAAFFGMATKYAEGLLAIKYRTKDTN 129
           AL TALAAT+GTGNIVGVATA+++GGPGA+FWMW+ A  GMATKYAE LLA+KYR +D N
Sbjct: 65  ALCTALAATIGTGNIVGVATAVQAGGPGAIFWMWLVALLGMATKYAECLLAVKYRVRDKN 124

Query: 130 GEISGGPMYYIINGMGQKWKPLAVFFSAAGILVALLGIGTFTQVNAIASSLEHTFKISTR 189
           G ++GGPMYYI  G+G +W  LA  F+  G++VA  GIGTF QVNAI  +++ TF I
Sbjct: 125 GFMAGGPMYYIERGLGIRW--LAKLFALFGVMVAFFGIGTFPQVNAITHAMQDTFNIPVL 182

Query: 190 FTSLILAVIVLFIIFGGIKSISKVSEKIVPFMAISYILATLIIIAVNYNKIPHTFQLIFS 249
            T++I+ ++V  II GG+K I+  S  IVPFMAI Y+  +L+II +N  K+P    LI
Sbjct: 183 VTAIIVTLLVGLIILGGVKRIATASSVIVPFMAILYVTTSLVIILLNIEKVPDAILLIID 242

Query: 250 GAFSGTAAIGGFSGAIVKEAIQKGIARGVFSNESGLGSAPIAAAAAKTKEPVEQGLISMT 309
            AF   AA+GG  G  V +AIQ G+ARG+FSNESGLGSAPIAAAAA+T+EPV QGLISMT
Sbjct: 243 SAFDPQAALGGAVGLTVMKAIQSGVARGIFSNESGLGSAPIAAAAAQTREPVRQGLISMT 302

Query: 310 GTFIDTIVICTLTGIAILVTGKWLEFDLQGAPLTQASFNTVFG-SLGSFALTFCLVLFAF 368
           GTF+DTI++CT+TGI +++TG W   +L GA +T  +F    G S+G+  +T  L+ FAF
Sbjct: 303 GTFLDTIIVCTMTGIVLVLTGAWNNPELAGATVTNYAFAQGLGTSIGATIVTVGLLFFAF 362

Query: 369 TTILGWSYYGERCFEYLFGTKFINAYRIIFVIMVGLGGFLQLDLIWVIADIVNGLMALPN 428
           TTILGW YYGERCF YL G + +  YR+ ++++VGLG FL L+LIW+IADIVNGLMA PN
Sbjct: 363 TTILGWCYYGERCFVYLVGIRGVKLYRLAYIMLVGLGAFLHLNLIWIIADIVNGLMAFPN 422

Query: 429 LIALLALSPIIVKETQKYFSETK 451
           LIAL+ L  +I++ET+ YF   K
Sbjct: 423 LIALIGLRKVIIEETKDYFQRLK 445
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5437> which encodes the amino acid sequence <SEQ ID 5438>. Analysis of this protein sequence reveals the following:

```
     Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.36   Transmembrane  183 - 199 ( 175 - 206)
   INTEGRAL    Likelihood = -7.80   Transmembrane  143 - 159 ( 140 - 163)
   INTEGRAL    Likelihood = -7.11   Transmembrane  209 - 225 ( 208 - 229)
   INTEGRAL    Likelihood = -5.95   Transmembrane  416 - 432 ( 413 - 434)
   INTEGRAL    Likelihood = -5.15   Transmembrane  304 - 320 ( 302 - 324)
   INTEGRAL    Likelihood = -4.46   Transmembrane  387 - 403 ( 382 - 408)
   INTEGRAL    Likelihood = -3.35   Transmembrane  348 - 364 ( 345 - 366)
   INTEGRAL    Likelihood = -1.17   Transmembrane   11 -  27 (  10 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF94579 GB:AE004221 sodium/alanine symporter [Vibrio cholerae]
 Identities = 261/441 (59%), Positives = 328/441 (74%), Gaps = 7/441 (1%)

Query: 3   ALVKLIDNLVWGPPLLILLVGTGIYLTSHLGLIQILKLPRAFKLIFSDDEG---HGDISS 59
           + ++ +D+LVWGPPLLILLVGTG+Y T  LGL+Q  +LP A  ++F  ++     GD+SS
Sbjct: 6   SFLQTVDSLVWGPPLLILLVGTGVYFTFRLGLLQFRRLPTALAMVFGREKSSDKQGDVSS 65

Query: 60  FAALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAAFFGMATKYAEGVLAIKYRTKD 119
           FAAL TAL+AT+GTGNIVGVATAIK GGPGALFWMW+AA FGMATKYAE +LA+KYR  D
Sbjct: 66  FAALCTALSATIGTGNIVGVATAIKLGGPGALFWMWLAALFGMATKYAECLLAVKYRQID 125

Query: 120 ANGHISGGPMYYIVNGMGTKWKPLAVLFAGSGILVALFGIGTFAQVNSITSSLGHSFGLS 179
             G + GGPMYY+ +G+ +K   LAVLFA   + VA FGIGTF QVN+I  +   SFG+
Sbjct: 126 DKGQMVGGPMYYLRDGVSSK--TLAVLFAVFAVGVACFGIGTFPQVNAILDATQISFGVP 183

Query: 180 PQMVSIVLAIFVAAIIFGGIHSISKVAEKVVPFMAIFYILSSLAVIFSHYQQLLPVIRLV 239
            +  ++VL + VA +  GGI SI+KVA KVVP MA+FYI++ L+VI ++   +L   + LV
Sbjct: 184 REASAVVLTVLVAIVTIGGIQSIAKVAGKVVPAMALFYIIACLSVIVTNADKLADAVELV 243

Query: 240 FQSAFTPTAAIGGFAGSLMKDAIQKGIARGVFSNESGLRSAPIAAAAAKTNEPVEQGLIS 299
             SAFT TAA GGF G+ +  AIQ GIARGVFSNESGL SAP+AAAAAKT+  VEQGLIS
Sbjct: 244 LVSAFTSTAATGGFLGASIMLAIQSGIARGVFSNESGLGSAPMAAAAAKTDSCVEQGLIS 303

Query: 300 MTGTFIDTIIICTLTGLSILVTGQWTGQLEGAPLTQSAFATVFG--NLGTFGLTFSLVLF 357
           MTGTF DTIIICT+TGL++++TG W   L GA +T  AFAT      +G   ++  L+ F
Sbjct: 304 MTGTFFDTIIICTMTGLALILTGAWQSDLSGAAMTTYAFATGLNAQTIGPMLVSIGLMFF 363

Query: 358 AFTTILGWSYYGERCFEFLFGITHLTYFRIVFILMVGLGGFLKLELIWVLADIVNGLMAL 417
           AFTTILGW+YYGERC  FLFG   +  ++IVFI ++  G FL L+LIW++ADIVNGLMA+
Sbjct: 364 AFTTILGWNYYGERCMVFLFGTKAVLPYKIVFIGLIASGAFLHLDLIWIIADIVNGLMAI 423

Query: 418 PNLIALLALSPVVILETKHYF 438
           PNLI L+AL  VV+ ETK YF
Sbjct: 424 PNLIGLVALRHVVVEETKQYF 444
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 323/439 (73%), Positives = 380/439 (85%), Gaps = 1/439 (0%)

Query: 9   MLTLFTHINSFVWGPPLLALLVGTGIYLSFRLGFIQLRQLSRAFKLIFREDNGQGDISSY 68
           M+ L   I++ VWGPPLL LLVGTGIYL+  LG IQ+ +L RAFKLIF +D G GDISS+
Sbjct: 1   MIALVKLIDNLVWGPPLLILLVGTGIYLTSHLGLIQILKLPRAFKLIFSDDEGHGDISSF 60

Query: 69  AALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAAFFGMATKYAEGLLAIKYRTKDT 128
           AALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAAFFGMATKYAEG+LAIKYRTKD
Sbjct: 61  AALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAAFFGMATKYAEGVLAIKYRTKDA 120

Query: 129 NGEISGGPMYYIINGMGQKWKPLAVFFSAAGILVALLGIGTFTQVNAIASSLEHTFKIST 188
           NG ISGGPMYYI+NGMG KWKPLAV F+ +GILVAL GIGTF QVN+I SSL H+F +S
Sbjct: 121 NGHISGGPMYYIVNGMGTKWKPLAVLFAGSGILVALFGIGTFAQVNSITSSLGHSFGLSP 180

Query: 189 RFTSLILAVIVLFIIFGGIKSISKVSEKIVPFMAISYILATLIIIAVNYNKIPHTFQLIF 248
           +  S++LA+ V  IIFGGI SISKV+EK+VPFMAI YIL++L +I  +Y ++     +L+F
Sbjct: 181 QMVSIVLAIFVAAIIFGGIHSISKVAEKVVPFMAIFYILSSLAVIFSHYQQLLPVIRLVF 240

Query: 249 SGAFSGTAAIGGFSGAIVKEAIQKGIARGVFSNESGLGSAPIAAAAAKTKEPVEQGLISM 308
             AF+ TAAIGGF+G+++K+AIQKGIARGVFSNESGL SAPIAAAAAKT EPVEQGLISM
Sbjct: 241 QSAFTPTAAIGGFAGSLMKDAIQKGIARGVFSNESGLRSAPIAAAAAKTNEPVEQGLISM 300

Query: 309 TGTFIDTIVICTLTGIAILVTGKWLEFDLQGAPLTQASFNTVFGSLGSFALTFCLVLFAF 368
           TGTFIDTI+ICTLTG++ILVTG+W    L+GAPLTQ++F TVFG+LG+F LTF LVLFAF
Sbjct: 301 TGTFIDTIIICTLTGLSILVTGQWTG-QLEGAPLTQSAFATVFGNLGTFGLTFSLVLFAF 359

Query: 369 TTILGWSYYGERCFEYLFGTKFINAYRIIFVIMVGLGGFLQLDLIWVIADIVNGLMALPN 428
           TTILGWSYYGERCFE+LFG   +  +RI+F++MVGLGGFL+L+LIWV+ADIVNGLMALPN
Sbjct: 360 TTILGWSYYGERCFEFLFGITHLTYFRIVFILMVGLGGFLKLELIWVLADIVNGLMALPN 419

Query: 429 LIALLALSPIIVKETQKYF 447
```

```
           LIALLALSP+++ ET+ YF
Sbjct: 420 LIALLALSPVVILETKHYF 438
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1751

A DNA sequence (GBSx1858) was identified in *S.agalactiae* <SEQ ID 5439> which encodes the amino acid sequence <SEQ ID 5440>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.16    Transmembrane   85 - 101 (  80 - 108)
   INTEGRAL    Likelihood = -5.36    Transmembrane  118 - 134 ( 115 - 137)
   INTEGRAL    Likelihood = -2.81    Transmembrane  177 - 193 ( 177 - 193)
   INTEGRAL    Likelihood = -0.48    Transmembrane   49 -  65 (  49 -  65)

----- Final Results -----
               bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12451 GB:Z99107 alternate gene name: ydxT~similar to cation
           efflux system membrane protein [Bacillus subtilis]
 Identities = 118/282 (41%), Positives = 181/282 (63%)

Query: 6   ENLQLAKRGPIISIIAYITLAVAKLAAGYWFDATSLVADGFNNLSDILGNVALLIGLHLA 65
           + L+  + G ++SI AY+ L+  KL  GY F + +L ADG NN +DI+ +VA+LIGL ++
Sbjct: 5   DELKKGESGALVSIAAYLVLSAIKLIIGYLFHSEALTADGLNNTTDIIASVAVLIGLRIS 64

Query: 66  SQPADSNHRFGHWKIEDLASLITSFIMFVVGIQVFIQTVTKIINNTDTNIDPLGAIVGAI 125
            +P D +H +GH++ E +ASLI SFIM VVG+QV       I +      D + A   A
Sbjct: 65  QKPPDEDHPYGHFRAETIASLIASFIMMVVGLQVLFSAGESIFSAKQETPDMIAAWTAAG 124

Query: 126 SALVMLGVYFYNKQLSQRVKSSALVAASKDNLSDAVTSIGTSIAIIAASLNFPIIDRLAA 185
            A++ML VY YNK+L+++VKS AL+AA+ DN SDA  SIGT I I+AA  +   ID + A
Sbjct: 125 GAVLMLIVYRYNKRLAKKVKSQALLAAAADNKSDAFVSIGTFIGIVAAQFHLAWIDTVTA 184

Query: 186 IIITYFILKTAYDIFIESAFSLSDGFDDYQLKQYEKAILTIPKISAVKSQRGRTYGSNIY 245
            +I   I KTA+DIF ES+ SL+DGFD   + Y++ I  I  +S +K  + R  GS ++
Sbjct: 185 FVIGLLICKTAWDIFKESSHSLTDGFDIKDISAYKQTIEKISGVSRLKDIKARYLGSTVH 244

Query: 246 LDIVLEMNPDLSVFESHAITERVEKLLSDKFSVYDIDIHVEP 287
           +D+V+E++ DL++ ESH I   +E+ + ++ ++    +H+EP
Sbjct: 245 VDVVVEVSADLNITESHDIANEIERRMKEEHAIDYSHVHMEP 286
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5441> which encodes the amino acid sequence <SEQ ID 5442>. Analysis of this protein sequence reveals the following:

```
     Possible site: 46
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.01    Transmembrane  121 - 137 ( 114 - 139)
   INTEGRAL    Likelihood = -5.41    Transmembrane   86 - 102 (  84 - 109)
   INTEGRAL    Likelihood = -5.04    Transmembrane  178 - 194 ( 176 - 197)
   INTEGRAL    Likelihood = -0.69    Transmembrane   50 -  66 (  50 -  66)
   INTEGRAL    Likelihood = -0.64    Transmembrane  158 - 174 ( 158 - 174)

----- Final Results -----
               bacterial membrane --- Certainty=0.4206(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12451 GB:Z99107 alternate gene name: ydxT~similar to cation
          efflux system membrane protein [Bacillus subtilis]
 Identities = 127/280 (45%), Positives = 187/280 (66%)

Query: 9   LKLARKGPIVSIIVYLSLSVAKLLAGYLLNASSLIADGFNNLSDIVGNVALLIGLHLASQ 68
            LK    G +VSI  YL LS  KL+ GYL ++ +L ADG NN +DI+ +VA+LIGL ++ +
Sbjct: 7   LKKGESGALVSIAAYLVLSAIKLIIGYLFHSEALTADGLNNTTDIIASVAVLIGLRISQK 66

Query: 69  PADANHKFGHWKIEDLSSLVTSFIMFLVGFQVLIHTIKSIFSGQQVDIDPLGAIVGIVSA 128
            P D +H +GH++ E ++SL+ SFIM +VG QVL    +SIFS +Q   D + A      A
Sbjct: 67  PPDEDHPYGHFRAETIASLIASFIMMVVGLQVLFSAGESIFSAKQETPDMIAAWTAAGGA 126

Query: 129 FVMLGVYVFNKRLSKRVKSSALVAASKDNLADAVTSIGTSIAIIAASLHLPVIDHIAAMI 188
             +ML VY +NKRL+K+VKS AL+AA+ DN +DA  SIGT I I+AA  HL  ID + A +
Sbjct: 127 VLMLIVYRYNKRLAKKVKSQALLAAAADNKSDAFVSIGTFIGIVAAQFHLAWIDTVTAFV 186

Query: 189 ITFFILKTAFDIFMESSFSLSDGFDSRHLKKYEKAILEIPKIVAVKSQRARTYGSNVYLD 248
            I   I KTA+DIF ESS SL+DGFD + +  Y++ I +I  +  +K  +AR  GS V++D
Sbjct: 187 IGLLICKTAWDIFKESSHSLTDGFDIKDISAYKQTIEKISGVSRLKDIKARYLGSTVHVD 246

Query: 249 IVLEMNPDLSVYESHSITEKVEQLLSDQFSIYDIDIHVEP 288
            +V+E++ DL++ ESH I  ++E+ + ++ +I    +H+EP
Sbjct: 247 VVVEVSADLNITESHDIANEIERRMKEEHAIDYSHVHMEP 286
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 274/406 (67%), Positives = 340/406 (83%), Gaps = 4/406 (0%)

Query: 7   NLQLAKRGPIISIIAYITLAVAKLAAGYWFDATSLVADGFNNLSDILGNVALLIGLHLAS 66
            NL+LA++GPI+SII Y++L+VAKL AGY  +A+SL+ADGFNNLSDI+GNVALLIGLHLAS
Sbjct: 8   NLKLARKGPIVSIIVYLSLSVAKLLAGYLLNASSLIADGFNNLSDIVGNVALLIGLHLAS 67

Query: 67  QPADSNHRFGHWKIEDLASLITSFIMFVVGIQVFIQTVTKIINNTDTNIDPLGAIVGAIS 126
            QPAD+NH+FGHWKIEDL+SL+TSFIMF+VG QV I T+  I +    +IDPLGAIVG +S
Sbjct: 68  QPADANHKFGHWKIEDLSSLVTSFIMFLVGFQVLIHTIKSIFSGQQVDIDPLGAIVGIVS 127

Query: 127 ALVMLGVYFYNKQLSQRVKSSALVAASKDNLSDAVTSIGTSIAIIAASLNFPIIDRLAAI 186
            A VMLGVY +NK+LS+RVKSSALVAASKDNL+DAVTSIGTSIAIIAASL+ P+ID +AA+
Sbjct: 128 AFVMLGVYVFNKRLSKRVKSSALVAASKDNLADAVTSIGTSIAIIAASLHLPVIDHIAAM 187

Query: 187 IITYFILKTAYDIFIESAFSLSDGFDDYQLKQYEKAILTIPKISAVKSQRGRTYGSNIYL 246
            IIT+FILKTA+DIF+ES+FSLSDGFD   LK+YEKAIL IPKI AVKSQR RTYGSN+YL
Sbjct: 188 IITFFILKTAFDIFMESSFSLSDGFDSRHLKKYEKAILEIPKIVAVKSQRARTYGSNVYL 247

Query: 247 DIVLEMNPDLSVFESHAITERVEKLLSDKFSVYDIDIHVEPASIPEDEIFDNVYQKLYKN 306
            DIVLEMNPDLSV+ESH+ITE+VE+LLSD+FS+YDIDIHVEPA IPE+EIFDNV +KLY+
Sbjct: 248 DIVLEMNPDLSVYESHSITEKVEQLLSDQFSIYDIDIHVEPAMIPEEEIFDNVAKKLYRY 307

Query: 307 EKIILAKIPGYETFISPDFYMINEKGNIITSDMLTNATNHSLASNFKYFNVKSISQKTKL 366
            EK+IL+K+P Y+ +I+  F +I+  G  +  +   N   +   SNF +F ++SISQKT L
Sbjct: 308 EKLILSKVPDYDHYIAKSFQLIDANGQTVNYEQFLNQEIY-YPSNFNHFQIESISQKTML 366

Query: 367 VSYELEGKRHTSIWRRNEKWFLIYHQIT--AKSSPYKTRRYQITSL 410
            V+Y+L G + TSIWRR+E W L++HQIT  AK   + T  Y+I  +
Sbjct: 367 VTYQLNGNQRTSIWRRHESWSLLFHQITPIAKKQLHHT-HYRIVKM 411
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1752

A DNA sequence (GBSx1859) was identified in *S.agalactiae* <SEQ ID 5443> which encodes the amino acid sequence <SEQ ID 5444>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12   Transmembrane  171 - 187 ( 161 - 194)
   INTEGRAL    Likelihood = -6.32   Transmembrane  118 - 134 ( 113 - 138)
   INTEGRAL    Likelihood = -5.89   Transmembrane   59 -  75 (  53 -  77)
   INTEGRAL    Likelihood = -5.52   Transmembrane  231 - 247 ( 226 - 252)
   INTEGRAL    Likelihood = -3.24   Transmembrane   86 - 102 (  84 - 103)
   INTEGRAL    Likelihood = -0.32   Transmembrane   31 -  47 (  31 -  47)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9869> which encodes amino acid sequence <SEQ ID 9870> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14850 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 80/226 (35%), Positives = 136/226 (59%), Gaps = 1/226 (0%)

Query: 27  TNNPIFGIMLTVWAYYIGIRIFRKYPSPAT-TPLLLATILLIAFLKLTHISYKDYYNGGS 85
            T +P FGI++++ A+ IG  +F+K       TPL +A +L IAFLK+   SY DY NGG
Sbjct: 4   TMSPYFGIVVSLAAFGIGTFLFKKTKGFFLFTPLFVAMVLGIAFLKIGGFSYADYNNGGE 63

Query: 86  FLTMLITPSTVVLAIPLYRTFHLMKHHIKSISISIILASVINTVFTAIVAKFFGMKYFLA 145
             +   + P+T+  AIPLY+    +K +   I  SII  S+ +     ++AK   +   +
Sbjct: 64  IIKFFLEPATIAFAIPLYKQRDKLKKYWWQIMASIIAGSICSVTIVYLLAKGIHLDSAVM 123

Query: 146 ISLFPKSVTTAMAVGITSKAGGLATITLVVVVITGILTSVLGPIFLKLLRIEDPVAIGLA 205
             S+ P++ TTA+A+ ++   GG++ IT   V+   ++   LG +FLK+ ++++P++ GLA
Sbjct: 124 KSMLPQAATTAIALPLSKGIGGISDITAFAVIFNAVIVYALGALFLKVFKVKNPISKGLA 183

Query: 206 LGGTGHAIGTGQALKYGQVQGAMAGLAIGITGICYVIVSPLVAGLI 251
            LG +GHA+G    ++ G+V+ AMA +A+ + G+  V+V P+   LI
Sbjct: 184 LGTSGHALGVAVGIEMGEVEAAMASIAVVVVGVVTVLVIPVFVQLI 229
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8893> and protein <SEQ ID 8894> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
SRCFLG: 0
McG: Length  of  UR:   22
     Peak Value of UR:   2.57
     Net Charge of CR: 0
McG: Discrim Score:      6.51
GvH: Signal Score (-7.5): -5.91
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 6 value:  -8.12 threshold:  0.0
   INTEGRAL    Likelihood = -8.12   Transmembrane  149 - 165 ( 139 - 172)
   INTEGRAL    Likelihood = -6.32   Transmembrane   96 - 112 (  91 - 116)
   INTEGRAL    Likelihood = -5.89   Transmembrane   37 -  53 (  31 -  55)
   INTEGRAL    Likelihood = -5.52   Transmembrane  209 - 225 ( 204 - 230)
   INTEGRAL    Likelihood = -3.24   Transmembrane   64 -  80 (  62 -  81)
   INTEGRAL    Likelihood = -0.32   Transmembrane    9 -  25 (   9 -  25)
   PERIPHERAL  Likelihood =  1.06     121
 modified ALOM score:   2.12
icm1 HYPID: 7  CFP: 0.425

*** Reasoning Step: 3
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01066(325 - 999 of 1305)
EGAD|107753|BS2884(4 - 229 of 231) hypothetical protein {Bacillus subtilis} OMNI|NT01BS3363
LrgB   GP|1770004|emb|CAA99613.1||Z75208   hypothetical   protein   {Bacillus   subtilis}
GP|2635355|emb|CAB14850.1||Z99118  similar  to  hypothetical  proteins  {Bacillus  subtilis}
PIR|D69983|D69983 conserved hypothetical protein ysbB - Bacillus subtilis
%Match = 17.2
%Identity = 35.4  %Similarity = 62.4
Matches = 80  Mismatches = 84  Conservative Sub.s = 61

192       222       252       282       312       342       372       402
WSTFKT*SPIFLG*LSLS*ERYFSIF*LLDWYPNGSKRDMKEIIQKLEVKMATLTNNPIFGIMLTVWAYYIGIRIFRKYP
                                                      | :| |||:::: |: ||  :|:|
                                                   MESTMSPYFGIVVSLAAFGIGTFLFKKTK
                                                           10        20

429       459       489       519       549       579       609       639
SPAT-TPLLLATILLIAFLKLTHISYKDYYNGGSFLTMLITPSTVVLAIPLYRTFHLMKHHIKSISISIILASVINTVFT
     |||::| :| |||||:   || || |||  :  :: |:|: :|||||:    :| :   |  |||  |: :
GFFLFTPLFVAMVLGIAFLKIGGFSYADYNNGGEIIKFFLEPATIAFAIPLYKQRDKLKKYWWQIMASIIAGSICSVTIV
         40        50        60        70        80        90       100

669       699       729       759       789       819       849       879
AIVAKFFGMKYFLAISLFPKSVTTAMAVGITSKAGGLATITLVVVVITGILTSVLGPIFLKLLRIEDPVAIGLALGGTGH
 ::||   :   :  |::|:: |||:|: ::   ||:: ||   |:  ::    || :|||::::::|:: ||||| :||
YLLAKGIHLDSAVMKSMLPQAATTAIALPLSKGIGGISDITAFAVIFNAVIVYALGALFLKVFKVKNPISKGLALGTSGH
        120       130       140       150       160       170       180

909       939       969       999      1029      1059      1089      1119
AIGTGQALKYGQVQGAMAGLAIGITGICYVIVSPLVAGLILK*G*GK*TQNNYVIIFKNRI*DK*L*YR*KK*LLERLSV
|:|    ::  |:|: ||| :|: : |:  |:| |:   ||
ALGVAVGIEMGEVEAAMASIAVVVVGVVTVLVIPVFVQLIGG
        200       210       220       230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1753

A DNA sequence (GBSx1860) was identified in *S.agalactiae* <SEQ ID 5445> which encodes the amino acid sequence <SEQ ID 5446>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA76857 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 44/194 (22%), Positives = 90/194 (45%), Gaps = 13/194 (6%)

Query: 21   TACSSSNTQQTSTSKSNVSQHKNIKADHEELRLKFNKVKLGVKANNFKGGTSLAELKQLF 80
            T  S ++T++ S+ K +  + K    D+ +L+  ++K+ +G   N+ +GG++  E+K +
Sbjct: 60   TNSSKNDTKKESSEKKSEDKSK----DNSDLKATYDKINVGDIMNSSEGGSTEDEVKAIL 115

Query: 81   GGEPNEKFDTPAGNVTLKGYRW-NVDD----ISITIQLLNDSSIVRSISNFKFIRDANIT 135
```

```
            GEP     T    ++     W NV       SIT+   +  +  +S+S  K  +   +T
Sbjct: 116 -GEPASSSTTDIQGISTTTLSWTNVKGGDLLASITVSFSDGKAASKSVSGLKVAKHDKVT 174

Query: 136 TKDYNSLKNGMSYN--KVKELLGEPDDISQAVSSDKEELQAAWISGIQSSDSDPGINLTF 193
               N++    SY+  + ++ LG+P  I+    + ++     W+  +   D    + ++F
Sbjct: 175 ADQVNNIATDGSYSEEQARKDLGDPTGITSTNINGEKNDTLIWMKNL-DGDLGATVTVSF 233

Query: 194 ENDKLTNKQQHGLK 207
            N    +K   GLK
Sbjct: 234 SNGNAISKSSSGLK 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5447> which encodes the amino acid sequence <SEQ ID 5448>. Analysis of this protein sequence reveals the following:

```
        Possible site: 21

   >>> May be a lipoprotein

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA76857 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 34/166 (20%), Positives = 74/166 (44%), Gaps = 8/166 (4%)

Query: 47  HQDKRANFEKIKLATVDSSFTGGTSLEELISLFGEPSQHDPKTAGEVTIDAYTWQFDQ-- 104
           + D +A ++KI +  + +S  GG++ +E+ ++ GEP+         ++    +W   +
Sbjct: 83  NSDLKATYDKINVGDIMNSSEGGSTEDEVKAILGEPASSSTTDIQGISTTTLSWTNVKGG 142

Query: 105 ---VTLTVNLYQNSSIVKTISNFTFARELGLSQKEYQQLQKGMSY--EDVKKILTEPDNY 159
               ++TV+     +  K++S    A+   ++  +   +    SY  E  +K L +P
Sbjct: 143 DLLASITVSFSDGKAASKSVSGLKVAKHDKVTADQVNNIATDGSYSEEQARKDLGDPTGI 202

Query: 160 SQASSSDHQTLQAIWVSGLKTDTSGANISLVFENNQLTEMSQVGLE 205
           +  + +  +    IW+  L  D  GA +++ F N      S  GL+
Sbjct: 203 TSTNINGEKNDTLIWMKNLDGDL-GATVTVSFSNGNAISKSSSGLK 247
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 84/199 (42%), Positives = 126/199 (63%), Gaps = 3/199 (1%)

Query: 11  TIVCLSFLG--LTACSSSNTQQTSTSKSNVSQHKNIKADHEELRLKFNKVKLGVKANNFK 68
           T++ +SF    L ACS++  ++   S S +   +  +A H++ R  F K+KL    ++F
Sbjct: 8   TLLLISFFTSFLVACSTTKDKEPQPSDSEIITPRLHQAAHQDKRANFEKIKLATVDSSFT 67

Query: 69  GGTSLAELKQLFGGEPNEKFDTPAGNVTLKGYRWNVDDISITIQLLNDSSIVRSISNFKF 128
           GGTSL EL  LFG EP++     AG VT+  Y W  D +++T+ L  +SSIV++ISNF F
Sbjct: 68  GGTSLEELISLFG-EPSQHDPKTAGEVTIDAYTWQFDQVTLTVNLYQNSSIVKTISNFTF 126

Query: 129 IRDANITTKDYNSLKNGMSYNKVKELLGEPDDISQAVSSDKEELQAAWISGIQSSDSDPG 188
            R+  ++ K+Y  L+ GMSY  VK++L EPD+ SQA SSD + LQA W+SG+++  S
Sbjct: 127 ARELGLSQKEYQQLQKGMSYEDVKKILTEPDNYSQASSSDHQTLQAIWVSGLKTDTSGAN 186

Query: 189 INLTFENDKLTNKQQHGLK 207
           I+L FEN++LT   Q GL+
Sbjct: 187 ISLVFENNQLTEMSQVGLE 205
```

SEQ ID 5446 (GBS650) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 178 (lane 9; MW 28kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1754

A DNA sequence (GBSx1861) was identified in *S.agalactiae* <SEQ ID 5449> which encodes the amino acid sequence <SEQ ID 5450>. This protein is predicted to be ribosomal protein S1 homolog; Sequence specific DNA-binding protein (r. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2950(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9363> which encodes amino acid sequence <SEQ ID 9364> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA97575 GB:U27517 ribosomal S1 protein [Homo sapiens]
 Identities = 156/305 (51%), Positives = 214/305 (70%), Gaps = 7/305 (2%)

Query: 1   MEARKAWDKLVGREGEVVTVKGTRAVKGGLSVEFEGLRGFIPASMIDTRFVRNTEKFVGQ 60
            ++ARKAW+ L   EG+ V  K   AV+GGL V+  G+RGF+PASM+  RFV +  +F  +
Sbjct: 53  LDARKAWENLSFAEGDTVDAKVINAVRGGLIVDVNGVRGFVPASMVAERFVSDLNQFKNK 112

Query: 61  EFDAKIKEVDAAENRFILSRREVVEESAAAARKEVFSNIEVGSVVTGKVARLTSFGAFID 120
            +  A++ E+D A  R ILSR+ V  +  AA   EVFS + VG VV G VARLT FGAF+D
Sbjct: 113 DIKAQVIEIDPANARLILSRKAVAAQERAAQLAEVFSKLSVGEVVEGTVARLTDFGAFVD 172

Query: 121 LGGVDGLVHVTELSHERNVSPKSVVTVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 180
            LGGVDGLVHV+E+SH+R  +P  V+T G++V+VK+L++D E GR+SLS+KAT  GPWD
Sbjct: 173 LGGVDGLVHVSEISHDRVKNPADVLTKGDKVDVKILALDTEKGRISLSIKATQRGPWDEA 232

Query: 181 EQKLAAGDVIEGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLSAGQEVTVK 240
              ++AAG V+EG VKR+ DFGAFVE+LPGI+GLVH+SQIS+KR+ENP +VL +G +V VK
Sbjct: 233 ADQIAAGSVLEGTVKRVKDFGAFVEILPGIEGLVHVSQISNKRIENPSEVLKSGDKVQVK 292

Query: 241 VLEVNSDAERVSLSMKALEERPAQAEGEKEEKRQSRPRRPRRQEKRDYELPETQTGFSMA 300
            VL++    ER+SLSMKALEE+P      + E R+       R +   Y+  +  +  ++
Sbjct: 293 VLDIKPAEERISLSMKALEEKP------EREDRRGNDGSASRADIAAYK-QQDDSAATLG 345

Query: 301 DLFGD 305
            D+FGD
Sbjct: 346 DIFGD 350
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5451> which encodes the amino acid sequence <SEQ ID 5452>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3312(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 284/309 (91%), Positives = 296/309 (94%), Gaps = 1/309 (0%)

Query: 1    MEARKAWDKLVGREGEVVTVKGTRAVKGGLSVEFEGLRGFIPASMIDTRFVRNTEKFVGQ 60
             +EARKAWDKLVGREGEVVTVKGTRAVKGGLSVEFEGLRGFIPASMIDTRFVRNTEKFVGQ
Sbjct: 93   LEARKAWDKLVGREGEVVTVKGTRAVKGGLSVEFEGLRGFIPASMIDTRFVRNTEKFVGQ 152
```

```
Query: 61  EFDAKIKEVDAAENRFILSRREVVEESAAAARKEVFSNIEVGSVVTGKVARLTSFGAFID 120
           EFDAKIKEVDAAENRFILSRREV+EE+A  AR EVFS I  G+VVTG VARLTSFGAFID
Sbjct: 153 EFDAKIKEVDAAENRFILSRREVIEEAAKEARAEVFSKISEGAVVTGTVARLTSFGAFID 212

Query: 121 LGGVDGLVHVTELSHERNVSPKSVVTVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 180
           LGGVDGLVHVTELSHERNVSPKSVV+VGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV
Sbjct: 213 LGGVDGLVHVTELSHERNVSPKSVVSVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 272

Query: 181 EQKLAAGDVIEGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLSAGQEVTVK 240
           EQKLA GDV+EGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLS GQEVTVK
Sbjct: 273 EQKLAQGDVVEGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLSVGQEVTVK 332

Query: 241 VLEVNSDAERVSLSMKALEERPAQAEGE-KEEKRQSRPRRPRRQEKRDYELPETQTGFSM 299
           VLEVN+  ERVSLS+KALEERPAQAEG+ KEEKRQSRPRRP+R+ +RDYELPETQTGFSM
Sbjct: 333 VLEVNAADERVSLSIKALEERPAQAEGDNKEEKRQSRPRRPKRESRRDYELPETQTGFSM 392

Query: 300 ADLFGDIEL 308
           ADLFGDIEL
Sbjct: 393 ADLFGDIEL 401
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1755

A DNA sequence (GBSx1862) was identified in *S.agalactiae* <SEQ ID 5453> which encodes the amino acid sequence <SEQ ID 5454>. This protein is predicted to be dihydroorotate dehydrogenase a (pyrD). Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1708(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB51330 GB:AJ131985 dihydroorotate dehydrogenase [Streptococcus pneumoniae]
 Identities = 227/310 (73%), Positives = 268/310 (86%)

Query: 1   MVSLKTEIAGFSFDNCLMNAAGIYCMTKEELLAIENSEAGSFVTKTGTLEAREGNPQPRY 60
           MVS KT+IAGF FDNCLMNAAG+ CMT EEL  ++NS AG+FVTKT TL+ R+GNP+PRY
Sbjct: 1   MVSTKTQIAGFEFDNCLMNAAGVACMTIEELEEVKNSAAGTFVTKTATLDFRQGNPEPRY 60

Query: 61  ADTDWGSINSMGLPNKGIDYYLDFVTELQDQDNSKNHVLSLVGLSPEETHIILKKVENSS 120
            D   GSINSMGLPN G+DYYLD++ +LQ++++++   LSLVG+SPEETH ILKKV+ S
Sbjct: 61  QDVPLGSINSMGLPNNGLDYYLDYLLDLQEKESNRTFFLSLVGMSPEETHTILKKVQESD 120

Query: 121 YNGLIELNLSCPNVPGKPQIAYDFEMTDLILSEIFSYYQKPLGIKLPPYFDIVHFDQAAT 180
           + GL ELNLSCPNVPGKPQIAYDFE TD IL+E+F+Y+ KPLGIKLPPYFDIV+FDQAA
Sbjct: 121 FRGLTELNLSCPNVPGKPQIAYDFETTDRILAEVFAYFTKPLGIKLPPYFDIVYFDQAAA 180

Query: 181 IFNKYPLAFINCVNSIGNGLVIDDETVVIKPKNGFGGIGGDFIKPTALANVHAFYKRLNP 240
           IFNKYPL F+NCVNSIGNGL I+DE+VVI+PKNGFGGIGG++IKPTALANVHAFY+RLNP
Sbjct: 181 IFNKYPLKFVNCVNSIGNGLYIEDESVVIRPKNGFGGIGGEYIKPTALANVHAFYQRLNP 240

Query: 241 SIKIIGTGGVKNGRDAFEHILCGASMVQIGTALQKEGPEIFQRVSRELKEIMADKGYQSL 300
            I+IIGTGGV  GRDAFEHILCGASMVQ+GT L KEG   F R++ ELK IM +KGY+SL
Sbjct: 241 QIQIIGTGGVLTGRDAFEHILCGASMVQVGTTLHKEGVSAFDRITNELKAIMVEKGYESL 300

Query: 301 EDFRGQLNYL 310
           EDFRG+L Y+
Sbjct: 301 EDFRGKLRYI 310
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5455> which encodes the amino acid sequence <SEQ ID 5456>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2689(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 239/309 (77%), Positives = 262/309 (84%)

Query: 1   MVSLKTEIAGFSFDNCLMNAAGIYCMTKEELLAIENSEAGSFVTKTGTLEAREGNPQPRY 60
           MVS  T+I  FSFDNCLMNAAG+YCMTKEEL+ +E S+A SFVTKTGTLE R GNP+PRY
Sbjct: 5   MVSTATQIGHFSFDNCLMNAAGVYCMTKEELMEVEKSQAASFVTKTGTLEVRPGNPEPRY 64

Query: 61  ADTDWGSINSMGLPNKGIDYYLDFVTELQDQDNSKNHVLSLVGLSPEETHIILKKVENSS 120
           ADT  GSINSMGLPN G  YYLDFV++L      K H LS+VGLSP ET  ILK +  S
Sbjct: 65  ADTRLGSINSMGLPNNGFRYYLDFVSDLAKTGQHKPHFLSVVGLSPTETETILKAIMASD 124

Query: 121 YNGLIELNLSCPNVPGKPQIAYDFEMTDLILSEIFSYYQKPLGIKLPPYFDIVHFDQAAT 180
           Y GL+ELNLSCPNVPGKPQIAYDFE TD +L  IF+YY KPLGIKLPPYFDIVHFDQAA
Sbjct: 125 YEGLVELNLSCPNVPGKPQIAYDFETTDQLLENIFTYYTKPLGIKLPPYFDIVHFDQAAA 184

Query: 181 IFNKYPLAFINCVNSIGNGLVIDDETVVIKPKNGFGGIGGDFIKPTALANVHAFYKRLNP 240
           IFNKYPL+F+NCVNSIGNGLVI DE V+IKPKNGFGGIGGD+IKPTALANVHAFYKRL P
Sbjct: 185 IFNKYPLSFVNCVNSIGNGLVIKDEQVLIKPKNGFGGIGGDYIKPTALANVHAFYKRLKP 244

Query: 241 SIKIIGTGGVKNGRDAFEHILCGASMVQIGTALQKEGPEIFQRVSRELKEIMADKGYQSL 300
           SI IIGTGGVK GRDAFEHILCGASMVQIGTAL +EGP IF+RV++ELK IM +KGYQSL
Sbjct: 245 SIHIIGTGGVKTGRDAFEHILCGASMVQIGTALHQEGPAIFERVTKELKTIMVEKGYQSL 304

Query: 301 EDFRGQLNY 309
           +DFRG L Y
Sbjct: 305 DDFRGNLRY 313
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1756

A DNA sequence (GBSx1863) was identified in *S.agalactiae* <SEQ ID 5457> which encodes the amino acid sequence <SEQ ID 5458>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4437(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89121 GB:AJ277485 beta-lactam resistance factor
           [Streptococcus pneumoniae]
 Identities = 238/410 (58%), Positives = 304/410 (74%)

Query: 1   MALKELTAKEFESYSGNYDLQSFMQTPEMAKLLKKRGYDITYMGYQIDGKMEIISIVYTI 60
```

```
           MAL  LT +EF++YS     +SFMQ+ +M  LL+KRG  I Y+  + +G++++ ++VY++
Sbjct: 1   MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60

Query: 61  PMTGGLHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDTYQEFTGEGKPKG 120
           PM GGLHME+NSGP ++    L  FY EL+ YAK  G LELL+KPY+TYQ F  +G P
Sbjct: 61  PMLGGLHMELNSGPIYTQQDALPVFYAELKEYAKQNGVLELLVKPYETYQTFDSQGNPID 120

Query: 121 APNTYLIDDLTSIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRPLVKKA 180
           A    +I DLT +GY  DGL  GYPGGEPDW Y K+L  +T ++LLKSFSKKG+PLVKKA
Sbjct: 121 AEKKSIIQDLTDLGYQFDGLTTGYPGGEPDWLYYKDLTELTEKSLLKSFSKKGKPLVKKA 180

Query: 181 MSFGIKIRVLKREELHIFKDITSSTSDRRDYMDKSLDYYQDFYDSFGDKAEFVIATLNFR 240
            +FGI+++ LKREEL IFK+IT  TS+RR+Y DKSL+YY+ FYD+FG++AEF+IA+LNF
Sbjct: 181 ETFGIRLKKLKREELSIFKNITKETSERREYSDKSLEYYEHFYDTFGEQAEFLIASLNFS 240

Query: 241 EYDHNLQLNAKKLEEQITVLDNRHQNNTDSAKYHRQRTELVNQLASLDKRRKEVEPFIQK 300
           +Y   LQ    KLEE +  L      N  S K   Q  E  +Q  + + R+ E    I+K
Sbjct: 241 DYMSKLQGEQSKLEENLDKLRLDLSKNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300

Query: 301 FGNQDVVLAGSLFIYSPKETVYLFSGSYTEFNKFYAPAVLQEYVMQEALKRQSTFYNFLG 360
           +G +D+VLAGSLF+Y P+ET YLFSGSYTEFNKFYAPA+LQ+YVM E++KR    YNFLG
Sbjct: 301 YGEEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLG 360

Query: 361 IQGNFDGSDGVLRFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRR 410
           IQG FDGSDGVLRFKQNFNGYIVRK GTFRY+P+PLKYK+IQLLKKI+ R
Sbjct: 361 IQGIFDGSDGVLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5459> which encodes the amino acid sequence <SEQ ID 5460>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 216/410 (52%), Positives = 291/410 (70%)

Query: 1   MALKELTAKEFESYSGNYDLQSFMQTPEMAKLLKKRGYDITYMGYQIDGKMEIISIVYTI 60
           MAL E++ ++F+ Y  +     SF+QT EMA L+ KRG    ++G + DG++++ ++V++
Sbjct: 1   MALIEISQEQFDHYCHSLVHHSFIQTSEMASLMAKRGAKPQFLGLEKDGELKVAAMVFSQ 60

Query: 61  PMTGGLHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDTYQEFTGEGKPKG 120
            + GG  ME+N+GP  ++ + L+HFY +L++YAK +  +EL++KPYD YQ F  +G P
Sbjct: 61  KVAGGWRMELNAGPNTNHPEELEHFYTQLKDYAKQKDVIELILKPYDNYQSFDTDGIPIS 120

Query: 121 APNTYLIDDLTSIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRPLVKKA 180
            PNT LI  LT++GY HDGL  GYP GEP WHYVK LEGI    L +SFSKKG+ L+KKA
Sbjct: 121 RPNTDLISLLTALGYKHDGLKTGYPEGEPVWHYVKKLEGIDSSRLTRSFSKKGKALIKKA 180

Query: 181 MSFGIKIRVLKREELHIFKDITSSTSDRRDYMDKSLDYYQDFYDSFGDKAEFVIATLNFR 240
            +FGIK+R LKR+ELH FK+IT +TSDRRDY+DKSL YYQDFYDSFGD  EF++ATLNF
Sbjct: 181 NTFGIKLRQLKRDELHHFKEITEATSDRRDYLDKSLSYYQDFYDSFGDSCEFMVATLNFE 240

Query: 241 EYDHNLQLNAKKLEEQITVLDNRHQNNTDSAKYHRQRTELVNQLASLDKRRKEVEPFIQK 300
           +Y +NL+    +L   I  +      N  S K   +  EL +Q  +   R  E   F+++
Sbjct: 241 DYLNNLKQRQLQLATSINKVKGDLGKNPHSEKKQNRLKELSSQFETFQVRISEALHFLEE 300

Query: 301 FGNQDVVLAGSLFIYSPKETVYLFSGSYTEFNKFYAPAVLQEYVMQEALKRQSTFYNFLG 360
           +G +DV LAGSLFIY+ +E VYLFSGSY +FNKFY+PA+LQE+ M +A+ +    YNFLG
Sbjct: 301 YGTKDVFLAGSLFIYTEQEAVYLFSGSYPKFNKFYSPALLQEHAMLKAIHKGIKQYNFLG 360

Query: 361 IQGNFDGSDGVLRFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRR 410
```

```
           I G FDGSDGVLRFKQNFNG+I++K GTFR YP P+KY  I+L KK+L R
Sbjct: 361 ITGKFDGSDGVLRFKQNFNGFILQKPGTFRCYPFPIKYHFIRLAKKLLNR 410
```

A related GBS gene <SEQ ID 8895> and protein <SEQ ID 8896> were also identified. Analysis of this protein sequence reveals the following:

```
Homology to resistance proteins
```

The protein has homology with the following sequences in the databases:

```
57.4/74.9% over 409aa
                                                                Streptococcus
pneumoniae
  GP|7649683| beta-lactam resistance factor Insert characterized

ORF01118(301 - 1530 of 1833)
GP|7649683|emb|CAB89121.1||AJ277485(1 - 410 of 410) beta-lactam resistance factor
{Streptococcus pneumoniae}
%Match = 39.0
%Identity = 57.3  %Similarity = 74.9
Matches = 235  Mismatches = 103  Conservative Sub.s = 72

240       270       300       330       360       390       420       450
IPVNRLLYKASNYVYALRKKRNS*LGKDTFMALKELTAKEFESYSGNYDLQSFMQTPEMAKLLKKRGYDITYMGYQIDGK
                              |||  || :||::||     :||||: :|  ||:|||  | |:  : :|:
                              MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGE
                                      10        20        30        40        50

480       510       540       570       600       630       660       690
MEIISIVYTIPMTGGLHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDTYQEFTGEGKPKGAPNTYLIDDL
::: ::||::|| ||||||:|||| ::    |  || ||: |||  | ||||:|||:||| |  :| |  |    :| ||
IQVAALVYSLPMLGGLHMELNSGPIYTQQDALPVFYAELKEYAKQNGVLELLVKPYETYQTFDSQGNPIDAEKKSIIQDL
        60        70        80        90       100       110       120       130

720       750       780       810       840       870       900       930
TSIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRPLVKKAMSFGIKIRVLKREELHIFKDITSSTSDRRD
| :||: |||  ||||||||| | |:|  :| ::|||||||||:|||||| :|||::: |||||| |||:||  ||:||:
TDLGYQFDGLTTGYPGGEPDWLYYKDLTELTEKSLLKSFSKKGKPLVKKAETFGIRLKKLKREELSIFKNITKETSERRE
       140       150       160       170       180       190       200       210

960       990       1020      1050      1080      1110      1140      1170
YMDKSLDYYQDFYDSFGDKAEFVIATLNFREYDHNLQLNAKKLEEQITVLDNRHQNNTDSAKYHRQRTELVNQLASLDKR
| ||||:||: |||:||::|||:||:||| :|   ||    |||| :  |      |  | | : |  |  :|: ::: |
YSDKSLEYYEHFYDTFGEQAEFLIASLNFSDYMSKLQGEQSKLEENLDKLRLDLSKNPHSEKKQNQLREYSSQFETFEVR
       220       230       240       250       260       270       280       290

1200      1230      1260      1290      1320      1350      1380      1410
RKEVEPFIQKFGNQDVVLAGSLFIYSPKETVYLFSGSYTEFNKFYAPAVLQEYVMQEALKRQSTFYXFLGIQGNFDGS G
: |   :|:|:| :|:|||||||:| |:|| |||||||||||||||||:||:||| |::||    | |||||| ||||:|
KAEARDLIEKYGEEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLGIQGIFDGSDG
       300       310       320       330       340       350       360       370

1440      1470      1500      1530      1560      1590      1620      1650
VLXFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRRT*KISLHKLIFYAL*KASFISLLLLFIQTIMFVI*RNFIT
|| ||||||||||||| |||||:|:|||||:|||||||: |
VLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR
       380       390       400       410
```

SEQ ID 8896 (GBS198) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 6; MW 48.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 6; MW 73.8kDa).

GBS198-GST was purified as shown in Figure 223, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1757

A DNA sequence (GBSx1864) was identified in *S.agalactiae* <SEQ ID 5461> which encodes the amino acid sequence <SEQ ID 5462>. This protein is predicted to be MurM protein. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4418(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89539 GB:AJ250767 MurM protein [Streptococcus pneumoniae]
 Identities = 204/410 (49%), Positives = 286/410 (69%), Gaps = 17/410 (4%)

Query: 1   MYRE---ITAVEHDRFVSESNQTNLLQSSNWPKVKDNWGSQLLGFFDGETQIASASILIK 57
            MYR    I  +E+D+FV E    N+LQSS W KVK +W  + LG ++GE  +A AS+LIK
Sbjct: 1   MYRYQIGIPTLEYDQFVKEHELANVLQSSAWEKVKSDWNHERLGVYEGENLLAVASVLIK 60

Query: 58  SLPLGFSMLYIPRGPIMDYSNLDIVTKVLKDLKAFGKKQRALFIKCDPLIYLK--MVNAK 115
            SLPLG+ M YIPRGPI+DY + +++  VL+ +K++ + +RA+F+  DP I L   +VN
Sbjct: 61  SLPLGYKMFYIPRGPILDYMDKELLKFVLQSIKSYARSKRAVFVTFDPSICLSQHLVN-- 118

Query: 116 DFENSPDEKEGLIAIDHLQRAGADWTGRTTDLAHTIQPRFQANLYANQFGLDKMSKKTRQ 175
              ++  +  E L  ++ L + G  W+G+TT++  TIQPR QA +Y   F  DK+SK TRQ
Sbjct: 119 --QDKREYPENLAIVEILGQLGVKWSGQTTEMDDTIQPRIQAKIYKENFEEDKLSKSTRQ 176

Query: 176 AIRTSKNKGVDIQFGSHELLEDFAELMKKTEDRKGINLRGIDYYQKLLDTYPNNSYITMA 235
            AIRT++NKG++IQ+G  ELL+ F+ELMKKTE RK I+LR   YY+KLLD +  +SYIT+
Sbjct: 177 AIRTARNKGLEIQYGGLELLDSFSELMKKTEKRKEIHLRNEAYYRKLLDNFKEDSYITLT 236

Query: 236 SLDVAKRLEKIEKECQIAQSERIKS--LELNREKKVKQHQGTIDRLNKEIDFLKEAQKAY 293
            +LDV+KRL ++E+  Q+A+++ ++    E  R  KV+  +   +RL +EIDFL +
Sbjct: 237 NLDVSKRLRELEE--QLAKNKALEEAFTESTRTSKVEAQKKEKERLVEEIDFL-QGYMNM 293

Query: 294 DRDIIPLAATLTLEFGNTSENIYAGMDDYFKSYSAPIYTWFETAQRAFERGNIWQNMGGI 353
            ++  IPLAATL+LEFG TS N+YAGMDD FK Y+API TW+ETA+ AFERG +WQN+GG+
Sbjct: 294 EKSNIPLAATLSLEFGTTSVNLYAGMDDDFKRYNAPILTWYETARYAFERGMVWQNLGGV 353

Query: 354 ENDLSGGLYHFKSKFEPIIEEFIGEFNIPVN---RLLYKASNYVYALRKK 400
            EN L+GGLYHFK KF P IEE++GEF +P +    LL  A ++   LRKK
Sbjct: 354 ENSLNGGLYHFKEKFNPTIEEYLGEFTMPTHPLYPLLRLALDFRKTLRKK 403
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5463> which encodes the amino acid sequence <SEQ ID 5464>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2239(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 203/399 (50%), Positives = 274/399 (67%), Gaps = 4/399 (1%)
```

```
Query: 5   ITAVEHDRFVSESNQTNLLQSSNWPKVKDNWGSQLLGFFDGETQIASASILIKSLPLGFS 64
           I+  EHD+FV    Q  LLQSS W KVKDNW  + + F++   Q+A+A+ LI+ LPLGF+
Sbjct: 13  ISPEEHDQFVLAQPQAGLLQSSKWGKVKDNWKHERISFYENGVQVAAAACLIRKLPLGFT 72

Query: 65  MLYIPRGPIMDYSNLDIVTKVLKDLKAFGKKQRALFIKCDPLIYLKMVNAKDFENSPDEK 124
           M+YIPRGPIMDY+N +++  V+K LK FGK +RALFIK DP + +K     + + S +
Sbjct: 73  MIYIPRGPIMDYANFELLDFVIKTLKTFGKSKRALFIKIDPSLVIKQT--LEGKESKEND 130

Query: 125 EGLIAIDHLQRAGADWTGRTTDLAHTIQPRFQANLYANQFGLDKMSKKTRQAIRTSKNKG 184
             L  I  L++ G +W+GRT +L  TIQPR QAN+YA  F  D + KK +Q+IRT+ NKG
Sbjct: 131 VTLSIIAFLKKLGVEWSGRTKELEDTIQPRIQANIYAKDFDFDSLPKKAKQSIRTATNKG 190

Query: 185 VDIQFGSHELLEDFAELMKKTEDRKGINLRGIDYYQKLLDTYPNNSYITMASLDVAKRLE 244
           V++  G  ELL+DF+ LMKKTE+RKGI LRG  YYQKLL  Y   SYITMASLD+ ++ +
Sbjct: 191 VNVTIGGSELLDDFSALMKKTENRKGIILRGKSYYQKLLGIYAGQSYITMASLDLPEQKK 250

Query: 245 KIEKECQIAQSERIKSLELNREKKVKQHQGTIDRLNKEIDFLKEAQKAYDRDIIPLAATL 304
            + ++   A +E+ +  + ++  KV ++Q TI RL K++  L E Q A  +  IPLAATL
Sbjct: 251 LLIQQLDKALAEQARLTDKSKPSKVAENQKTIARLQKDLTILSE-QLATGQTRIPLAATL 309

Query: 305 TLEFGNTSENIYAGMDDYFKSYSAPIYTWFETAQRAFERGNIWQNMGGIENDLSGGLYHF 364
           TL +G TSEN+YAGMDD +++Y AP+ TW+ETA+ AF+RG  W N+GG+EN   GGLYHF
Sbjct: 310 TLIYGETSENLYAGMDDDYRNYQAPLLTWYETAKEAFKRGCRWHNLGGVENQQDGGLYHF 369

Query: 365 KSKFEPIIEEFIGEFNIPVNRLLYKASNYVYALRKKRNS 403
           K++  P IEEF GEFNIPV  L+   +   Y LRKK  S
Sbjct: 370 KARLNPTIEEFAGEFNIPVG-LVSSLAILTYNLRKKLRS 407
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1758

A DNA sequence (GBSx1865) was identified in *S.agalactiae* <SEQ ID 5465> which encodes the amino acid sequence <SEQ ID 5466>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1759

A DNA sequence (GBSx1866) was identified in *S.agalactiae* <SEQ ID 5467> which encodes the amino acid sequence <SEQ ID 5468>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.07   Transmembrane   56 -  72 (  55 -  74)

----- Final Results -----
```

```
                 bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9625> which encodes amino acid sequence <SEQ ID 9626> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89120 GB:AJ277484 beta-lactam resistance factor
           [Streptococcus pneumoniae]
 Identities = 166/410 (40%), Positives = 250/410 (60%), Gaps = 10/410 (2%)

Query: 6   MYHVTVGISEKEYDAFAIASSQTNLLHSSKWAQVKSNWQNERLGFYKDDQLVAVASILIK 65
           MY   +GI   EYD F       N+L SS W +VKSNWQ+E+ G Y++++L+A ASILI+
Sbjct: 1   MYRYQIGIPTLEYDQFVKEHELANVLQSSAWEEVKSNWQHEKFGVYREEKLLATASILIR 60

Query: 66  SLPLGFTMLYIPRGPIMDYSNKELVNFVLKTLKNFGRKKRAVFAKFDPALLLRQYHLKEE 125
           +LPLG+ M YIPRGPI+DY +KEL+NF ++++K++ R KRAVF  FDP++ L Q  + +E
Sbjct: 61  TLPLGYKMFYIPRGPILDYGDKELLNFAIQSIKSYARSKRAVFVTFDPSICLSQSLINQE 120

Query: 126 NVAEEIDESRQAIDNLKSAGAQWIGPTKAISETIQPRFQANIYTKANIEENFPKHTKRLI 185
               E  E+   ID+L+  G +W G T+ + +TIQPR QA IY +    E+   K TK+ I
Sbjct: 121 KT--EFPENLAIIDSLQQMGVRWSGKTEEMGDTIQPRIQAKIYKENFEEDKLSKSTKQAI 178

Query: 186 KDAKHRGVQIYRANIDDLPKFATVVALTENRKGVALRNENYFHQLMTIYGEDAYLYLAKV 245
           + A+++G++I    ++ L  F+ ++  TE RK + LRNE Y+ +L+  + + AY+ LA +
Sbjct: 179 RTARNKGLEIQYGGLELLDSFSELMKKTEKRKEIHLRNEAYYKKLLDNFKDKAYITLATL 238

Query: 246 NLPKRLAQFKEQLLQIQKDLSETPSHQKSRLTRLNQQEASVKQYILEFQEFSKKYPD--- 302
           ++ KR  + +EQL +  + L ET + + +R +++  Q+   K+ +LE   F ++Y D
Sbjct: 239 DVSKRSQELEEQLAK-NRALEETFT-ESTRTSKVEAQKKE-KERLLEELTFLQEYIDVGQ 295

Query: 303 -EPVIAGILSIRFGNVLEMLYAGMDDSFRKFYPQYLLNARVFEDAFKNDIVSANLGGVEG 361
               +A  LS+ FG     +YAGMDD F+++    L        AF+  ++  NLGGVE
Sbjct: 296 ARVPLAATLSLEFGTTSVNIYAGMDDDFKRYNAPILTWYETARYAFERGMIWQNLGGVEN 355

Query: 362 SLNDGLTKFKSNFNPMFEEYIGEFNLAINPLLYKLANLAYTIRKKQRHSH 411
           SLN GL  FK  FNP  EEY+GEF +  +P LY L  LA   RK  R  H
Sbjct: 356 SLNGGLYHFKEKFNPTIEEYLGEFTMPTHP-LYPLLRLALDFRKTLRKKH 404
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5469> which encodes the amino acid sequence <SEQ ID 5470>. Analysis of this protein sequence reveals the following:

```
      Possible site: 32

   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.32    Transmembrane    59 -   75 (   59 -   75)

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB89120 GB:AJ277484 beta-lactam resistance factor
           [Streptococcus pneumoniae]
 Identities = 166/402 (41%), Positives = 255/402 (63%), Gaps = 5/402 (1%)

Query: 9   KIGISEEEHDSFVKEHQQISVLQGSDWAKIKNQWQNERIGIYKEEKQVASLSLLIKLLPL 68
           +IGI   E+D FVKEH+  +VLQ S W ++K+ WQ+E+ G+Y+EEK +A+ S+LI+ LPL
Sbjct: 5   QIGIPTLEYDQFVKEHELANVLQSSAWEEVKSNWQHEKFGVYREEKLLATASILIRTLPL 64

Query: 69  GRSIIYIPRGPVMDYLDRDLVAFTMKTLKDYGKTKKALFIKYDPAILLKQYALGQEEEEK 128
           G  + YIPRGP++DY D++L+ F ++++K Y ++K+A+F+ +DP+I L Q  + QE+ E
```

```
Sbjct: 65  GYKMFYIPRGPILDYGDKELLNFAIQSIKSYARSKRAVFVTFDPSICLSQSLINQEKTEF 124

Query: 129 PLALAAIKNLQEAGVHWTGLTMEIADSIQPRFQANIYTQENLEMQFPKHTRRLIKDAKQR 188
           P  LA I +LQ+ GV W+G T E+ D+IQPR QA IY +   E +  K T++ I+ A+ +
Sbjct: 125 PENLAIIDSLQQMGVRWSGKTEEMGDTIQPRIQAKIYKENFEEDKLSKSTKQAIRTARNK 184

Query: 189 GVKTYRVSQSELHKFSKIVSLTEKRKNISLRNEAYFQKLMTTYGDKAYLHLAKVNIPQKL 248
           G++        L  FS+++  TEKRK I LRNEAY++KL+  + DKAY+ LA +++ ++
Sbjct: 185 GLEIQYGGLELLDSFSELMKKTEKRKEIHLRNEAYYKKLLDNFKDKAYITLATLDVSKRS 244

Query: 249 DQYRQQLILINQDITRTQAHQKKRLKKLEDQKASLERYITE---FEGFTDQYPEEVVVAG 305
            +  +QL   N+ +  T   +  R  K+E QK   ER + E    + + D     V +A
Sbjct: 245 QELEEQLAK-NRALEETFT-ESTRTSKVEAQKKEKERLLEELTFLQEYIDVGQARVPLAA 302

Query: 306 ILSISYGNVMEMLYAGMNDDFKKFYPQYLLYPNVFQDAYQDGIIWANMGGVEGSLDDGLT 365
            LS+ +G     +YAGM+DDFK++    L +    + A++ G+IW N+GGVE SL+ GL
Sbjct: 303 TLSLEFGTTSVNIYAGMDDDFKRYNAPILTWYETARYAFERGMIWQNLGGVENSLNGGLY 362

Query: 366 KFKANFAPTIEEFIGEFNLPVSPLYHIANTMYKIRKQLKNKH 407
            FK  F PTIEE++GEF +P  PLY +       RK L+ KH
Sbjct: 363 HFKEKFNPTIEEYLGEFTMPTHPLYPLLRLALDFRKTLRKKH 404
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 226/407 (55%), Positives = 318/407 (77%), Gaps = 3/407 (0%)

Query: 5   LMYHVTVGISEKEYDAFAIASSQTNLLHSSKWAQVKSNWQNERLGFYKDDQLVAVASILI 64
           L ++  +GISE+E+D+F     Q ++L  S WA++K+ WQNER+G YK+++ VA  S+LI
Sbjct: 4   LTFYAKIGISEEEHDSFVKEHQQISVLQGSDWAKIKNQWQNERIGIYKEEKQVASLSLLI 63

Query: 65  KSLPLGFTMLYIPRGPIMDYSNKELVNFVLKTLKNFGRKKRAVFAKFDPALLLRQYHLKE 124
           K LPLG +++YIPRGP+MDY +++LV F +KTLK++G+ K+A+F K+DPA+LL+QY L +
Sbjct: 64  KLLPLGRSIIYIPRGPVMDYLDRDLVAFTMKTLKDYGKTKKALFIKYDPAILLKQYALGQ 123

Query: 125 ENVAEEIDESRQAIDNLKSAGAQWIGPTKAISETIQPRFQANIYTKANIEENFPKHTKRL 184
           E   EE   +  AI NL+ AG  W G T  I+++IQPRFQANIYT+ N+E  FPKHT+RL
Sbjct: 124 EE--EEKPLALAAIKNLQEAGVHWTGLTMEIADSIQPRFQANIYTQENLEMQFPKHTRRL 181

Query: 185 IKDAKHRGVQIYRANIDDLPKFATVVALTENRKGVALRNENYFHQLMTIYGEDAYLYLAK 244
           IKDAK RGV+ YR +  +L KF+ +V+LTE RK ++LRNE YF +LMT YG+ AYL+LAK
Sbjct: 182 IKDAKQRGVKTYRVSQSELHKFSKIVSLTEKRKNISLRNEAYFQKLMTTYGDKAYLHLAK 241

Query: 245 VNLPKRLAQFKEQLLQIQKDLSETPSHQKSRLTRLNQQEASVKQYILEFQEFSKKYPDEP 304
           VN+P++L Q+++QL+ I +D++ T +HQK RL +L  Q+AS+++YI EF+ F+ +YP+E
Sbjct: 242 VNIPQKLDQYRQQLILINQDITRTQAHQKKRLKKLEDQKASLERYITEFEGFTDQYPEEV 301

Query: 305 VIAGILSIRFGNVLEMLYAGMDDSFRKFYPQYLLNARVFEDAFKNDIVSANLGGVEGSLN 364
           V+AGILSI +GNV+EMLYAGM+D F+KFYPQYLL   VF+DA+++ I+ AN+GGVEGSL+
Sbjct: 302 VVAGILSISYGNVMEMLYAGMNDDFKKFYPQYLLYPNVFQDAYQDGIIWANMGGVEGSLD 361

Query: 365 DGLTKFKSNFNPMFEEYIGEFNLAINPLLYKLANLAYTIRKKQRHSH 411
           DGLTKFK+NF P  EE+IGEFNL ++P LY +AN  Y IRK+ ++ H
Sbjct: 362 DGLTKFKANFAPTIEEFIGEFNLPVSP-LYHIANTMYKIRKQLKNKH 407
```

SEQ ID 5468 (GBS377) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 4; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 4; MW 74kDa).

GBS377-GST was purified as shown in Figure 212, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1760

A DNA sequence (GBSx1867) was identified in *S.agalactiae* <SEQ ID 5471> which encodes the amino acid sequence <SEQ ID 5472>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2073(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9627> which encodes amino acid sequence <SEQ ID 9628> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76720 GB:AE000446 orf, hypothetical protein [Escherichia coli K12]
 Identities = 127/269 (47%), Positives = 189/269 (70%), Gaps = 1/269 (0%)

Query: 7   SIKLVAVDIDGTLLNSKREITPEVAKAVQEAKSKGVKIVIATGRPIIGVQDLLEELKLNE 66
            +IKL+A+D+DGTLL     I+P V  A+  A+++GV +V+ TGRP  GV + L+EL + +
Sbjct: 2   AIKLIAIDMDGTLLLPDHTISPAVKNAIAAARARGVNVVLTTGRPYAGVHNYLKELHMEQ 61

Query: 67  EGDYVITFNGGLVQDTATGDDIIKETLTYEDYLDFELLARKLGVHMHAITKEGIYTANRD 126
             GDY IT+NG LVQ  A G  + +  L+Y+DY   E L+R++G H HA+ +  +YTANRD
Sbjct: 62  PGDYCITYNGALVQKAADGSTVAQTALSYDDYRFLEKLSREVGSHFHALDRTTLYTANRD 121

Query: 127 IGKYTIHEVTLVNMPLFYRTPEEMG-DKEIIKLMMIDQPDILDAAIAKIPKKVLDNYTIV 185
            I  YT+HE  +  +PL +   E+M  + + +K+MMID+P ILD AIA+IP++V + YT++
Sbjct: 122 ISYYTVHESFVATIPLVFCEAEKMDPNTQFLKVMMIDEPAILDQAIARIPQEVKEKYTVL 181

Query: 186 KSTPFYLEILPKNVNKGTALLHLAEKMGLTVDQTMAIGDEENDRAMLEVVGNPVVMQNGN 245
            KS P++LEIL K VNKGT +  LA+ +G+  ++ MAIGD+END AM+E  G  V M N
Sbjct: 182 KSAPYFLEILDKRVNKGTGVKSLADVLGIKPEEIMAIGDQENDIAMIEYAGVGVAMDNAI 241

Query: 246 PELKKIAKYITKSNEESGVAYALREWVIN 274
            P +K++A ++TKSN E GVA+A+ ++V+N
Sbjct: 242 PSVKEVANFVTKSNLEDGVAFAIEKYVLN 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3407> which encodes the amino acid sequence <SEQ ID 3408>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3474(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 197/268 (73%), Positives = 235/268 (87%)

Query: 7   SIKLVAVDIDGTLLNSKREITPEVAKAVQEAKSKGVKIVIATGRPIIGVQDLLEELKLNE 66
            SIKLVAVDIDGTLL   R IT +V +AVQEAK++GV +VIATGRPI GV  LLE+L+LN
Sbjct: 2   SIKLVAVDIDGTLLTDDRRITDDVFQAVQEAKAQGVHVVIATGRPIAGVISLLEQLELNH 61

Query: 67  EGDYVITFNGGLVQDTATGDDIIKETLTYEDYLDFELLARKLGVHMHAITKEGIYTANRD 126
            +G++VITFNGGLVQD  TG++I+KE +TY+DYL+ E L+RKLGVHMHAITKEGIYTANR+
Sbjct: 62  KGNHVITFNGGLVQDAETGEEIVKELMTYDDYLETEFLSRKLGVHMHAITKEGIYTANRN 121

Query: 127 IGKYTIHEVTLVNMPLFYRTPEEMGDKEIIKLMMIDQPDILDAAIAKIPKKVLDNYTIVK 186
```

```
           IGKYT+HE TLVNMP+FYRTPEEM +KEIIK+MMID+PD+LDAAI +IP+   D YTIVK
Sbjct: 122 IGKYTVHESTLVNMPIFYRTPEEMTNKEIIKMMMIDEPDLLDAAIKQIPQHFFDKYTIVK 181

Query: 187 STPFYLEILPKNVNKGTALLHLAEKMGLTVDQTMAIGDEENDRAMLEVVGNPVVMQNGNP 246
           STPFYLE +PK V+KG A+ HLA+K+GL + QTMAIGD ENDRAMLEVV NPVVM+NG P
Sbjct: 182 STPFYLEFMPKTVSKGNAIKHLAKKLGLDMSQTMAIGDAENDRAMLEVVANPVVMENGVP 241

Query: 247 ELKKIAKYITKSNEESGVAYALREWVIN 274
           ELKKIAKYITKSN +SGVA+A+R+WV+N
Sbjct: 242 ELKKIAKYITKSNNDSGVAHAIRKWVLN 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1761

A DNA sequence (GBSx1868) was identified in *S.agalactiae* <SEQ ID 5473> which encodes the amino acid sequence <SEQ ID 5474>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2360(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07537 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 211/423 (49%), Positives = 285/423 (66%), Gaps = 5/423 (1%)

Query: 3   EKVFRDPVHTYIHVNNQVIYDLINTKEFQRLRRIKQTSTTSFTFHGAEHSRFSHCLGVYE 62
           EKVF+DPVH YIHV +++I+ LI TKEFQRLRR++Q  TT  TFHGAEH+RF+H LGVYE
Sbjct: 12  EKVFKDPVHRYIHVRDELIWALIGTKEFQRLRRVRQLGTTFLTFHGAEHTRFNHSLGVYE 71

Query: 63  LARKVTEIFDEHYSDLWNKNESLLTMAAALLHDIGHGAYSHTFERLFNTDHEAYTQEIIT 122
           + R++ E+F       WN+ E LLT+ AALLHDIGHG +SH+FE++F+TDHE +T+ +I
Sbjct: 72  ITRRIIEVFQGR--PYWNEEERLLTLCAALLHDIGHGPFSHSFEKVFDTDHEEWTRRMIV 129

Query: 123 NPTTEINAILRKVAPDFPDKVASVINHSYPNKQVVQLISSQIDCDRMDYLLRDSYYTAAS 182
             T EI+ +L K+  DFP KVA VI  +YPNK V  +ISSQID DRMDYL RD+YYT  S
Sbjct: 130 GDT-EIHNVLLKMGDDFPQKVADVIEKTYPNKLVTSIISSQIDADRMDYLQRDAYYTGVS 188

Query: 183 YGQFDLTRILRVIRPTDSGIAFARNGMHAVEDYIVSRFQMYMQVYFHPASRAMELLLQNL 242
           YG FD+ RILRV+RP +  +   ++GMHAVEDYI+SR+QMY QVYFHP +R+ E++L  +
Sbjct: 189 YGHFDMERILRVMRPMEDQVVIKQSGMHAVEDYIMSRYQMYWQVYFHPVTRSAEVILSKV 248

Query: 243 LKRARFLFDTHRDFFEQTSPNLIPFFTDQYDLQDYLALDDGVMNTYFQSWMQADDNILAD 302
            KR + L++     F+Q   +    F     L DYL LD+ +   YFQ W + +D IL+D
Sbjct: 249 FKRVKDLYEQGYK-FKQEPKHFYSLFEGNMSLDDYLRLDESITMYYFQIWQEEEDRILSD 307

Query: 303 LANRFINRKVFKSITFEESDKEN-LVKMKELVSQVGFDPDYYTGVHANFDLPYDVYRPEH 361
           L  RFINR++FK I F  + + N   ++++L +Q   DP+YY  V ++ DLPYD YRP
Sbjct: 308 LCVRFINRQLFKYIEFNPNLQMNDWPRLQQLFAQAEIDPEYYLVVDSSSDLPYDFYRPGE 367

Query: 362 SNPRTEIQIIQKNGQLAELSSLSPIVKALTGSNYGDQRFYFPKEMLTLDSLFSSTKEEFQ 421
              R  I +I  NG+L ELS  S +V+A++G    D + YFP + LT  S     K+E
Sbjct: 368 EEERLPIHLIMPNGKLRELSRESDVVEAISGKKRTDHKLYFPMDCLTDQSDHKEIKQEIL 427

Query: 422 SYI 424
           S +
Sbjct: 428 SLL 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5475> which encodes the amino acid sequence <SEQ ID 5476>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2220(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 321/428 (75%), Positives = 379/428 (88%)

Query: 1   MNEKVFRDPVHTYIHVNNQVIYDLINTKEFQRLRRIKQTSTTSFTFHGAEHSRFSHCLGV 60
           MNEKVFRDPVH YIH++N +IYDLINTKEFQRLRRIKQ  TT+FTFHGAEHSRFSHCLGV
Sbjct: 1   MNEKVFRDPVHNYIHIDNPLIYDLINTKEFQRLRRIKQVPTTAFTFHGAEHSRFSHCLGV 60

Query: 61  YELARKVTEIFDEHYSDLWNKNESLLTMAAALLHDIGHGAYSHTFERLFNTDHEAYTQEI 120
           YE+AR+VT IF+E Y+D+WNK+ESL+TM AALLHDIGHGAYSHTFE LF+TDHEA+TQEI
Sbjct: 61  YEIARRVTAIFEEKYADIWNKDESLVTMTAALLHDIGHGAYSHTFEVLFHTDHEAFTQEI 120

Query: 121 ITNPTTEINAILRKVAPDFPDKVASVINHSYPNKQVVQLISSQIDCDRMDYLLRDSYYTA 180
           ITNP TEINAIL + APDFPDKVASVINH+YPNKQVVQLISSQIDCDRMDYLLRDSY++A
Sbjct: 121 ITNPETEINAILVRHAPDFPDKVASVINHTYPNKQVVQLISSQIDCDRMDYLLRDSYFSA 180

Query: 181 ASYGQFDLTRILRVIRPTDSGIAFARNGMHAVEDYIVSRFQMYMQVYFHPASRAMELLLQ 240
           A+YGQFDL RILRVIRP + GI F  +GMHAVEDYIVSRFQMYMQVYFHPASRA+EL+LQ
Sbjct: 181 ANYGQFDLMRILRVIRPVEDGIVFEHSGMHAVEDYIVSRFQMYMQVYFHPASRAVELILQ 240

Query: 241 NLLKRARFLFDTHRDFFEQTSPNLIPFFTDQYDLQDYLALDDGVMNTYFQSWMQADDNIL 300
           NLLKRA+ L+   + +F++T+P LIPFF  + +L DY+ALDDGVMNTYFQ WM ++D+IL
Sbjct: 241 NLLKRAQHLYPEQQAYFQKTAPGLIPFFEKKANLADYIALDDGVMNTYFQVWMASEDHIL 300

Query: 301 ADLANRFINRKVFKSITFEESDKENLVKMKELVSQVGFDPDYYTGVHANFDLPYDVYRPE 360
           +DLA+RFINRK+ KS+TF++  +  L ++++LV  VGFDPDYYTG+H NFDLPYD+YRPE
Sbjct: 301 SDLASRFINRKILKSVTFDQDSQGELERLRQLVESVGFDPDYYTGIHINFDLPYDIYRPE 360

Query: 361 HSNPRTEIQIIQKNGQLAELSSLSPIVKALTGSNYGDQRFYFPKEMLTLDSLFSSTKEEF 420
             NPRT+I+++QK+G LAELS LSPIVKALTG+ YGD+RFYFPKEML LD LF+ +KE F
Sbjct: 361 LENPRTQIEMMQKDGSLAELSQLSPIVKALTGTTYGDRRFYFPKEMLELDDLFAPSKETF 420

Query: 421 QSYITNEH 428
            SYI+N H
Sbjct: 421 MSYISNGH 428
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1762

A DNA sequence (GBSx1869) was identified in *S.agalactiae* <SEQ ID 5477> which encodes the amino acid sequence <SEQ ID 5478>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4789(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5479> which encodes the amino acid sequence <SEQ ID 5480>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3650(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 64/127 (50%), Positives = 89/127 (69%)

Query: 5   MKLEINNNIQIDNETEMIHEIHDCQFIEKGSYVYLNYINAEGERVVIKANHEELLMTRFS 64
           MKL++ N+I+  +ETE+I EIHDC++ EKG Y YL Y N + E+VVIK N  EL M+RFS
Sbjct: 1   MKLQLTNHIRFGDETEIIQEIHDCEWREKGGYQYLIYQNTDKEKVVIKYNETELTMSRFS 60

Query: 65  NPKSVMRFHRETPALVNIPTPLGVQHLITETSHYQFDLSQQRLHINYVLKQTETGDCFAN 124
           NP+S+M+F      L+ +PTP+GVQ  +T+TSHY  D S Q+L ++Y L Q +T   FA+
Sbjct: 61  NPQSIMKFFAGKKVLIALPTPMGVQQFLTDTSHYHLDCSCQKLDLHYHLLQAQTEMLFAS 120

Query: 125 YELRIQW 131
           Y L + W
Sbjct: 121 YHLELSW 127
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1763

A DNA sequence (GBSx1870) was identified in *S.agalactiae* <SEQ ID 5481> which encodes the amino acid sequence <SEQ ID 5482>. This protein is predicted to be cation-transporting ATPase PacL (ctpF). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.27   Transmembrane  256 - 272 ( 246 - 276)
   INTEGRAL    Likelihood = -9.02   Transmembrane   64 -  80 (  58 -  85)
   INTEGRAL    Likelihood = -8.49   Transmembrane  833 - 849 ( 828 - 855)
   INTEGRAL    Likelihood = -8.17   Transmembrane   89 - 105 (  81 - 107)
   INTEGRAL    Likelihood = -7.48   Transmembrane  864 - 880 ( 860 - 884)
   INTEGRAL    Likelihood = -3.29   Transmembrane  287 - 303 ( 284 - 306)
   INTEGRAL    Likelihood = -2.55   Transmembrane  754 - 770 ( 753 - 773)
   INTEGRAL    Likelihood = -0.85   Transmembrane  695 - 711 ( 694 - 711)
   INTEGRAL    Likelihood = -0.75   Transmembrane  793 - 809 ( 792 - 809)

----- Final Results -----
               bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13439 GB:Z99112 similar to calcium-transporting ATPase
           [Bacillus subtilis]
 Identities = 380/888 (42%), Positives = 545/888 (60%), Gaps = 49/888 (5%)

Query: 10  FYTQGQEEVLTSLESS-REGLSTTEAKNRLEMYGRNELEEGKKRSLIAKFFDQFKDLMII 68
           F+  GQ ++L +  +S ++GL+  E K RL+ +G NEL+EGKK S +  FF QFKD M++
Sbjct: 3   FHEMGQTDLLEATNTSMKQGLTEKEVKKRLDKHGPNELQEGKKTSALLLFFAQFKDFMVL 62

Query: 69  ILLVAAALSVITEGMHG-LTDALIILAVVILNAAFGVYQEGQAEAAIEALKDMSSPIARV 127
```

```
           +LL A  +S    G  G   DA+ I+A+V +N   G +QE +AE +++ALK++S+P
Sbjct: 63  VLLAATLIS----GFLGEYVDAVAIIAIVFVNGILGFFQERRAEQSLQALKELSTPHVMA 118

Query: 128 RRDGHTIEVDSKELVPGDLVMLEAGDVVPADLRLLEAASLKIEEAALTGESVPVEKDISQ 187
            R+G   ++ SKELVPGD+V   +GD + AD+R++EA SL+IEE+ALTGES+PV K   +
Sbjct: 119 LREGSWTKIPSKELVPGDIVKFTSGDRIGADVRIVEARSLEIEESALTGESIPVVKHADK 178

Query: 188 VVAEDAGIGDRVNMAYQNSNVTYGRGYGVVTNTGMYTEVGKIADMLANADESETPLKQSL 247
           +   D  +GD  NMA+  + VT G G GVV  TGM T +GKIADML +A    TPL++ L
Sbjct: 179 LKKPDVSLGDITNMAFMGTIVTRGSGVGVVVGTGMNTAMGKIADMLESAGTLSTPLQRRL 238

Query: 248 VQLSKLLTYLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAIPEGLPAIVTIVLSMG 307
            QL K+L  + +++ V+   VG+ ++        +  V+LAVAAIPEGLPAIVT+ LS+G
Sbjct: 239 EQLGKILIVVALLLTVLVVAVGV-IQGHDLYSMFLAGVSLAVAAIPEGLPAIVTVALSLG 297

Query: 308 TKTLAKRNSIVRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYT-------------- 353
            + + K+ SIVRKLPAVETLG   II SDKTGT+T N+MTV  V++
Sbjct: 298 VQRMIKQKSIVRKLPAVETLGCASIICSDKTGTMTQNKMTVTHVWSGGKTWRVAGAGYEP 357

Query: 354 NGVLQSSSEEISVDNNTL--------RIMNFSNDTKIDPSGKLIGDPTETALVQFGLDKN 405
            G    + +EISV+ +           + N SN  K D    L GDPTE AL+
Sbjct: 358 KGSFTLNEKEISVNEHKPLQQMLLFGALCNNSNIEKRDGEYVLDGDPTEGALLTAARKGG 417

Query: 406 FDVREVLKNEPRVAELPFDSDRKLMSTIHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLV 465
           F    V  N   + E PFDS RK+M+ I +  D + +I  KGAPD L++R ++I  +G
Sbjct: 418 FSKEFVESNYRVIEEFPFDSARKMMTVIVENQDRKRYIITKGAPDVLMQRSSRIYYDGSA 477

Query: 466 RDITAEDKEAILNTNKELAKQALRVLMMAYK--YETQIPSLETDIVESDLVFSGLVGMID 523
              + E K       + LA QALR + +AY+     + PS+E    E DL   GL G+ID
Sbjct: 478 ALFSNERKAETEAVLRHLASQALRTIAVAYRPIKAGETPSMEQ--AEKDLTMLGLSGIID 535

Query: 524 PERPEAAEAVRVAKEAGIRPIMITGDHQDTAEAIAKRLGIIDANDTEDHVFTGAELNELS 583
           P RPE  +A++  +EAGI+ +MITGDH +TA+AIAK L ++  +     +  G  LNELS
Sbjct: 536 PPRPEVRQAIKECREAGIKTVMITGDHVETAKAIAKDLRLLPKS---GKIMDGKMLNELS 592

Query: 584 DEEFQKVFKQYSVYARVSPEHKVRIVKAWQNDGKVVAMTGDGVNDAPSLKTADIGIGMGI 643
            EE   V +   V+ARVSPEHK++IVKA+Q +G +VAMTGDGVNDAP++K ADIG+ MGI
Sbjct: 593 QEELSHVVEDVYVFARVSPEHKLKIVKAYQENGHIVAMTGDGVNDAPAIKQADIGVSMGI 652

Query: 644 TGTEVSKGASDMVLADDNFATIIVAVEEGRKVFSNIQKSIQYLLSANMAEVFTIFFATLL 703
           TGT+V+K AS +VL DDNFATI  A++EGR ++ NI+K I+YLL++N+ E+  + FA LL
Sbjct: 653 TGTDVAKEASSLVLVDDNFATIKSAIKEGRNIYENIRKFIRYLLASNVGEILVMLFAMLL 712

Query: 704 GWDV-LAPVHLLWINLVTDTLPAIALGVEPAEPGVMTHKPRGRQSNFFDGGVMGAIIYQG 762
              + L P+ +LW+NLVTD LPA+ALG++  E  VM  KPR  +   F   +   ++ +G
Sbjct: 713 ALPLPLVPIQILWVNLVTDGLPAMALGMDQPEGDVMKRKPRHPKEGVFARKLGWKVVSRG 772

Query: 763 ILQTILVLGVYGWALMY---PEHAGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIF 819
            L  I V  +  + ++Y   PE+  Y      A T+AFATL L QL+H F+ +S   S+F
Sbjct: 773 FL--IGVATILAFIIVYHRNPENLAY------AQTIAFATLVLAQLIHVFDCRS-ETSVF 823

Query: 820 TVGAFKNRTFNWSIPVAFILLMVTIVVPGFNKLFHVTHLSSTQWLTVV 867
           +   F+N     ++  + +L++V I  P    +FH   ++   W+ V+
Sbjct: 824 SRNPFQNLYLIGAVLSSILLMLVVIYYPPLQPIFHTVAITPGDWMLVI 871
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4171> which encodes the amino acid sequence <SEQ ID 4172>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.47   Transmembrane  863 - 879 ( 856 - 883)
   INTEGRAL    Likelihood =-10.08   Transmembrane   64 -  80 (  58 -  86)
   INTEGRAL    Likelihood = -8.97   Transmembrane  256 - 272 ( 249 - 275)
   INTEGRAL    Likelihood = -8.55   Transmembrane   89 - 105 (  81 - 107)
   INTEGRAL    Likelihood = -5.84   Transmembrane  832 - 848 ( 827 - 850)
   INTEGRAL    Likelihood = -3.13   Transmembrane  287 - 303 ( 284 - 307)
   INTEGRAL    Likelihood = -2.66   Transmembrane  762 - 778 ( 761 - 779)
```

```
INTEGRAL    Likelihood = -0.37   Transmembrane  685 - 701 ( 685 - 701)

----- Final Results -----
                  bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 735/892 (82%), Positives = 813/892 (90%), Gaps = 1/892 (0%)

Query: 3   KEQKKSLFYTQGQEEVLTSLESSREGLSTTEAKNRLEMYGRNELEEGKKRSLIAKFFDQF 62
           KEQ+   FYTQ +E VL  LE+SREGL++ +AK RL  YGRNEL+EG+KRSL  KF DQF
Sbjct: 3   KEQRHEAFYTQSEETVLAQLETSREGLTSAQAKERLAEYGRNELDEGEKRSLFMKFLDQF 62

Query: 63  KDLMIIILLVAAALSVITEGMHGLTDALIILAVVILNAAFGVYQEGQAEAAIEALKDMSS 122
           KDLMIIIL+VAA LSV+TEGM GLTDA+IILAVVILNAAFGVYQEGQAEAAIEALK MSS
Sbjct: 63  KDLMIIILIVAALLSVLTEGMEGLTDAIIILAVVILNAAFGVYQEGQAEAAIEALKSMSS 122

Query: 123 PIARVRRDGHTIEVDSKELVPGDLVMLEAGDVVPADLRLLEAASLKIEEAALTGESVPVE 182
           P+AR+RRDGH  E+DSKELVPGD+V+LEAGDVVPADLRLLEA SLKIEEAALTGESVPVE
Sbjct: 123 PLARIRRDGHVTEIDSKELVPGDIVLLEAGDVVPADLRLLEANSLKIEEAALTGESVPVE 182

Query: 183 KDISQVVAEDAGIGDRVNMAYQNSNVTYGRGYGVVTNTGMYTEVGKIADMLANADESETP 242
           KD+S  V+EDAGIGDRVNM YQNSNVTYGRG GV+TNTGMYTEVG IA MLANADE++TP
Sbjct: 183 KDLSTAVSEDAGIGDRVNMGYQNSNVTYGRGIGVITNTGMYTEVGHIAGMLANADETDTP 242

Query: 243 LKQSLVQLSKLLTYLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAIPEGLPAIVTI 302
           LKQ+L  LSK+LTY I++IA +TF VG+F+R +  +EGLMTSVALAVAAIPEGLPAIVT+
Sbjct: 243 LKQNLDNLSKILTYAILVIAAVTFAVGVFLRGQHPLEGLMTSVALAVAAIPEGLPAIVTV 302

Query: 303 VLSMGTKTLAKRNSIVRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYTNGVLQSSSE 362
           VLS+GT+ LAKRN+I+RKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYTNG LQSSS
Sbjct: 303 VLSLGTQVLAKRNAIIRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYTNGTLQSSSA 362

Query: 363 EISVDNNTLRIMNFSNDTKIDPSGKLIGDPTETALVQFGLDKNFDVREVLKNEPRVAELP 422
           +I+ DN TLR+MNF+NDTK+DPSGKLIGDPTETALV+FGLD NFDVRE +  EPRVAELP
Sbjct: 363 DIAFDNTTLRVMNFANDTKVDPSGKLIGDPTETALVEFGLDHNFDVREAMVAEPRVAELP 422

Query: 423 FDSDRKLMSTIHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLVRDITAEDKEAILNTNKE 482
           FDSDRKLMSTIHK++DG+YFIAVKGAPDQLLKRVT+IE+NG +R IT  DK+ IL+TNK
Sbjct: 423 FDSDRKLMSTIHKQADGKYFIAVKGAPDQLLKRVTQIEENGQIRPITDADKKTILDTNKS 482

Query: 483 LAKQALRVLMMAYKYETQIPSLETDIVESDLVFSGLVGMIDPERPEAAEAVRVAKEAGIR 542
           LAKQALRVLMMAYKY   +P+LET+IVE++LVFSGLVGMIDPERPEAA+AV+VAKEAGIR
Sbjct: 483 LAKQALRVLMMAYKYSDALPTLETEIVEANLVFSGLVGMIDPERPEAAQAVKVAKEAGIR 542

Query: 543 PIMITGDHQDTAEAIAKRLGIIDANDTEDHVFTGAELNELSDEEFQKVFKQYSVYARVSP 602
           PIMITGDHQDTA+AIAKRLGII+  D  DHVFTGAELNELSDEEFQKVFKQYSVYARVSP
Sbjct: 543 PIMITGDHQDTAKAIAKRLGIIE-EDGVDHVFTGAELNELSDEEFQKVFKQYSVYARVSP 601

Query: 603 EHKVRIVKAWQNDGKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNF 662
           EHKVRIVKAWQN+GKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNF
Sbjct: 602 EHKVRIVKAWQNEGKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNF 661

Query: 663 ATIIVAVEEGRKVFSNIQKSIQYLLSANMAEVFTIFFATLLGWDVLAPVHLLWINLVTDT 722
           ATIIVAVEEGRKVFSNIQK+IQYLLSANMAEVFTIF ATL GWDVL PVHLLWINLVTDT
Sbjct: 662 ATIIVAVEEGRKVFSNIQKTIQYLLSANMAEVFTIFLATLFGWDVLQPVHLLWINLVTDT 721

Query: 723 LPAIALGVEPAEPGVMTHKPRGRQSNFFDGGVMGAIIYQGILQTILVLGVYGWALMYPEH 782
           LPAIALGVEPAEPGVM HKPRGR+S+FFDGGV  AI+YQG  QTILVLGVYG+ALM+PEH
Sbjct: 722 LPAIALGVEPAEPGVMKHKPRGRKSSFFDGGVKEAILYQGAFQTILVLGVYGFALMFPEH 781

Query: 783 AGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIFTVGAFKNRTFNWSIPVAFILLMV 842
             Y  +HADALTMA+ TLGLIQLVHA+NVKSVYQSIFTVG FKN+ FN+SIPVAF+ LM
Sbjct: 782 TSYHDVHADALTMAYVTLGLIQLVHAYNVKSVYQSIFTVGLFKNKLFNYSIPVAFVALMA 841

Query: 843 TIVVPGFNKLFHVTHLSSTQWLTVVIGSLLMVVLTEIVKFIQRKLGQDEKAI 894
```

```
           T+VVPGFN+ FHVTHL+ TQWL V+IGSLLMVVL E+VK +QR LGQDEKAI
Sbjct: 842 TVVVPGFNQFFHVTHLTITQWLVVIIGSLLMVVLVELVKAVQRSLGQDEKAI 893
```

A related GBS gene <SEQ ID 8897> and protein <SEQ ID 8898> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     -9.88
GvH: Signal Score (-7.5): -6.96
     Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 9 value: -13.27 threshold:  0.0
   INTEGRAL    Likelihood =-13.27   Transmembrane  256 - 272 ( 246 - 276)
   INTEGRAL    Likelihood = -9.02   Transmembrane   64 -  80 (  58 -  85)
   INTEGRAL    Likelihood = -8.49   Transmembrane  833 - 849 ( 828 - 855)
   INTEGRAL    Likelihood = -8.17   Transmembrane   89 - 105 (  81 - 107)
   INTEGRAL    Likelihood = -7.48   Transmembrane  864 - 880 ( 860 - 884)
   INTEGRAL    Likelihood = -3.29   Transmembrane  287 - 303 ( 284 - 306)
   INTEGRAL    Likelihood = -2.55   Transmembrane  754 - 770 ( 753 - 773)
   INTEGRAL    Likelihood = -0.85   Transmembrane  695 - 711 ( 694 - 711)
   INTEGRAL    Likelihood = -0.75   Transmembrane  793 - 809 ( 792 - 809)
   PERIPHERAL  Likelihood =  1.06     714
 modified ALOM score:   3.15

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01112(328 - 2901 of 3282)
EGAD|108247|BS1566(3 - 871 of 890) hypothetical protein {Bacillus subtilis} OMNI|NT01BS1841
cation-transporting ATPase PacL GP|2337795|emb|CAA74269.1||Y13937 putative PacL protein
{Bacillus subtilis} GP|2633938|emb|CAB13439.1||Z99112 similar to calcium-transporting
ATPase {Bacillus subtilis} PIR|H69877|H69877 calcium-transporting ATPase homolog yloB -
Bacillus subtilis
%Match = 29.0
%Identity = 43.9  %Similarity = 64.5
Matches = 376  Mismatches = 291  Conservative Sub.s = 176

249       279       309       339       369       396       426       456
GVVLNSETCFHKNRSLFVCGETKGGKVLLKEQKKSLFYTQGQEEVLTSLESS-REGLSTTEAKNRLEMYGRNELEEGKKR
                                    |:  || ::| :  :| ::||:  | | ||: :| |||:||||
                                  MKFHEMGQTDLLEATNTSMKQGLTEKEVKKRLDKHGPNELQEGKKT
                                          10        20        30        40

486       516       546       576       606       636       666       696
SLIAKFFDQFKDLMIIILLVAAALSVITEGMHGLTDALIILAVVILNAAFGVYQEGQAEAAIEALKDMSSPIARVRRDGH
| :  || ||||:|:::||   | ::|:  :    ||: |:|:| :|  :| :|| :|| :::|||::|:|     |:|
SALLLFFAQFKDFMVLVLL---AATLISGFLGEYVDAVAIIAIVFVNGILGFFQERRAEQSLQALKELSTPHVMALREGS
            60           70        80        90       100       110       120

726       756       786       816       846       876       906       936
TIEVDSKELVPGDLVMLEAGDVVPADLRLLEAASLKIEEAALTGESVPVEKDISQVVAEDAGIGDRVNMAYQNSNVTYGR
  :: ||||||||:| : :|| : ||:|::|| ||:|||:||||||:|| |   ::   |  :||  |||:  : || |
WTKIPSKELVPGDIVKFTSGDRIGADVRIVEARSLEIEESALTGESIPVVKHADKLKKPDVSLGDITNMAFMGTIVTRGS
              140       150       160       170       180       190       200

966       996      1026      1056      1086      1116      1146      1176
GYGVVTNTGMYTEVGKIADMLANADESETPLKQSLVQLSKLLTYLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAI
| |||  ||| | :||||||| :|    |||:: | || |:|  : ::: |:   ||: ::       ::  |:||||||
GVGVVVGTGMNTAMGKIADMLESAGTLSTPLQRRLEQLGKILIVVALLLTVLVVAVGV-IQGHDLYSMFLAGVSLAVAAI
              220       230       240       250       260        270       280
```

```
1206      1236      1266      1296      1326      1356                1374
PEGLPAIVTIVLSMGTKTLAKRNSIVRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYT--------------NGVLQ
|||||||||: ||:| : : |: |||||||||||||   || ||||||:| |:|||  |::               | :
PEGLPAIVTVALSLGVQRMIKQKSIVRKLPAVETLGCASIICSDKTGTMTQNKMTVTHVWSGGKTWRVAGAGYEPKGSFT
               300       310       320       330       340       350       360

1404                1440      1470      1500      1530      1560      1590
SSSEEISVDNNT--------LRIMNFSNDTKIDPSGKLIGDPTETALVQFGLDKNFDVREVLKNEPRVAELPFDSDRKLM
 : :||||: :           : | ||  | |    | ||||| ||:       |    |  |   : |:|||| ||:|
LNEKEISVNEHKPLQQMLLFGALCNNSNIEKRDGEYVLDGDPTEGALLTAARKGGFSKEFVESNYRVIEEFPFDSARKMM
               380       390       400       410       420       430       440

1620      1650      1680      1710      1740      1770      1794      1824
STIHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLVRDITAEDKEAILNTNKELAKQALRVLMMAYK--YETQIPSLETDI
: | :  | : :|  ||||| |::| ::|  :|     : | |       : || |||| : :||:     : ||:|
TVIVENQDRKRYIITKGAPDVLMQRSSRIYYDGSAALFSNERKAETEAVLRHLASQALRTIAVAYRPIKAGETPSME--Q
               460       470       480       490       500       510         520

1854      1884      1914      1944      1974      2004      2034      2064
VESDLVFSGLVGMIDPERPEAAEAVRVAKEAGIRPIMITGDHQDTAEAIAKRLGIIDANDTEDHVFTGAELNELSDEEFQ
 | ||   || |:||| |||  :|::  :||||: :|||||| :||:|||| |           :  |  ||||| ||:
AEKDLTMLGLSGIIDPPRPEVRQAIKECREAGIKTVMITGDHVETAKAIAKDL---RLLPKSGKIMDGKMLNELSQEELS
       530       540       550       560       570          580       590

2094      2124      2154      2184      2214      2244      2274      2304
KVFKQYSVYARVSPEHKVRIVKAWQNDGKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNFATIIVA
 | :   |:||||||||::||||:| :| :||||||||||||::| ||||: ||||||:|:| || :|| |||||||  |
HVVEDVYVFARVSPEHKLKIVKAYQENGHIVAMTGDGVNDAPAIKQADIGVSMGITGTDVAKEASSLVLVDDNFATIKSA
          610       620       630       640       650       660       670

2334      2364      2394                2451      2481      2511      2541
VEEGRKVFSNIQKSIQYLLSANMAEVFTIFFATLLGWDV-LAPVHLLWINLVTDTLPAIALGVEPAEPGVMTHKPRGRQS
::||| :: ||:| |:|||::|: |:: ::|| ||   : | |:::||:||||| |||:|||::  |  ||  |||  :
IKEGRNIYENIRKFIRYLLASNVGEILVMLFAMLLALPLPLVPIQILWVNLVTDGLPAMALGMDQPEGDVMKRKPRHPKE
          690       700       710       720       730       740       750

2571      2601      2631      2661      2691      2721      2751      2781
NFFDGGVMGAIIYQGILQTILVLGVYGWALMYPEHAGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIFTVGAFKNR
  |   :   :: :| |    |  :  : ::|  |       | | |:||||| | ||:| |: :|   |:|:   |:|
GVFARKLGWKVVSRGFLIG--VATILAFIIVY--HRN-PENLAYAQTIAFATLVLAQLIHVFDCRS-ETSVFSRNPFQNL
          770         780         790        800       810        820       830

2811      2841      2871      2901      2931      2961      2991      3021
TFNWSIPVAFILLMVTIVVPGFNKLFHVTHLSSTQWLTVVIGSLLMVVLTEIVKFIQRKLGQDEKAI*FS**KNSLRISK
 :  ::  : :|::| |  | :  :||  ::    |: |:
YLIGAVLSSILLMLVVIYYPPLQPIFHTVAITPGDWMLVIGMSAIPTFLLAGSLLTRKK
                850       860       870       880       890
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1764

A DNA sequence (GBSx1871) was identified in *S.agalactiae* <SEQ ID 5483> which encodes the amino acid sequence <SEQ ID 5484>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2905(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB48940 GB:AJ248283 hypothetical protein [Pyrococcus abyssi]
 Identities = 60/221 (27%), Positives = 100/221 (45%), Gaps = 37/221 (16%)

Query: 33  KIDHLHIA------GDISNHFTKDTLP-FINNLKKH---IKLSYNLGNHDMLDLTE--TE 80
           KID L I       GD+SN+   D +   I+ L      + L    GNHD+  L +
Sbjct: 15  KIDVLKIPDIAIQLGDLSNYGEPDIIENLISELVTQLDPVPLLVIPGNHDIYGLNDIFAA 74

Query: 81  IQRLDFQTYR------------FDKKMLLAFHGWYDYSFSNN--RDIKDVEKLKKTFWFD 126
            QR +    R             ++  ++   GWYDYS +       KD  ++K  F F
Sbjct: 75  FQRFNKLVKRAGAIPLMEGPLILEEIGIVGVPGWYDYSLAPGYLNMTKDEYEIK-AFGFR 133

Query: 127 RR-----LKRPNNDVTIQASILKRLDEILAKVDSS--NIIIAMHFVPHKQFTMT--HPRF 177
           R      +K   +D  +    L  L++ ++++  S  ++I+A+HF P K       +P
Sbjct: 134 RLEDADYIKSSLSDEELVRWNLNLLEKFISEIRESVNDVILALHFAPFKDSLKYTGNPEI 193

Query: 178 SPFNAFLGSQAYHDLFQKYHIKDVVFGHAHRSFGDVKIGET 218
             F+A++GSQ + +   +++I  +V GH HRS  +  IG+T
Sbjct: 194 DYFSAYMGSQRFGEFALRHNIGLIVHGHTHRSI-EYYIGKT 233
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1765

A DNA sequence (GBSx1872) was identified in *S.agalactiae* <SEQ ID 5485> which encodes the amino acid sequence <SEQ ID 5486>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.18   Transmembrane  173 - 189 ( 173 - 189)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB16056 GB:Z99124 fructose-1,6-bisphosphatase [Bacillus subtilis]
 Identities = 314/642 (48%), Positives = 446/642 (68%), Gaps = 7/642 (1%)

Query: 2   SNFYKLLKEKFPRKEDIVTEMINLEAICQLPKGTEYFISDLHGEYDAVDYLLRTGAGSIR 61
           S +  LL +K+  +E +VTE+INL+AI  LPKGTE+F+SDLHGEY A  ++LR G+G ++
Sbjct: 33  SKYLDLLAQKYDCEEKVVTEIINLKAILNLPKGTEHFVSDLHGEYQAFQHVLRNGSGRVK 92

Query: 62  AKLLDCFDWQKIVAVDLDDFCILLYYPKEKLAFDKMNLSASAYKTKLW-EMIPLQIQVLK 120
            K+ D F    I   ++D+   L+YYP++KL   K +  A     + + E I   I+++
Sbjct: 93  EKIRDIFSGV-IYDREIDELAALVYYPEDKLKLIKHDFDAKEALNEWYKETIHRMIKLVS 151

Query: 121 YFSSKYTKSKVRKQLSGKFAYIIEELLAEIDRNPEKKSYFDTIIEKLFELDQVEDLIIVL 180
           Y SSKYT+SK+RK L  +FAYI EELL + ++   K+ Y+  II+++ EL Q + LI  L
Sbjct: 152 YCSSKYTRSKLRKALPAQFAYITEELLYKTEQAGNKEQYYSEIIDQIIELGQADKLITGL 211

Query: 181 SQTIQVLIIDHLHVVGDIYDRGRYPDRILNRLMAFPNLDIQWGNHDVTWMGAASGSYLCM 240
           + ++Q L++DHLHVVGDIYDRG  PDRI+  L+ + ++DIQWGNHDV W+GA SGS +C+
Sbjct: 212 AYSVQRLVVDHLHVVGDIYDRGPQPDRIMEELINYHSVDIQWGNHDVLWIGAYSGSKVCL 271

Query: 241 VNVIRIAARYNNITLIEDRYGINLRRLVDYSRRYYEPLPSFVPILDGEEMTHPDELDLLN 300
            N+IRI ARY+N+ +IED YGINLR L++ + +YY+  P+F P  D  E    DE+  +
Sbjct: 272 ANIIRICARYDNLDIIEDVYGINLRPLLNLAEKYYDDNPAFRPKAD--ENRPEDEIKQIT 329

Query: 301 MIQQATAILQFKLEAQLIDRRPEFQMHNRQLINQVNYKDLSISIKEVVHQLKDFNSRCID 360
            I QA A++QFKLE+ +I RRP F M  R L+ +++Y    I++    +QL++     I+
Sbjct: 330 KIHQAIAMIQFKLESPIIKRRPNFNMEERLLLEKIDYDKNEITLNGKTYQLENTCFATIN 389
```

```
Query: 361 SKNPSRLTSEEEELLQQLMIAFQTSESLKKHIDFLFEKGSMYLTYNDNLLFHGCIPMHSN 420
            + P +L  EE E++ +L+ + Q SE L +H++F+ +KGS+YL YN NLL HGCIP+  N
Sbjct: 390 PEQPDQLLEEEAEVIDKLLFSVQHSEKLGRHMNFMMKKGSLYLKYNGNLLIHGCIPVDEN 449

Query: 421 GDFKSFKIAGKTYGGRDLLDLFESQIRLAYARPEKHDDLATDIIWYLWCGENSSLFGKNA 480
           G+ ++  I  K Y GR+LLD+FE  +R A+A PE+ DDLATD+ WYLW GE SSLFGK A
Sbjct: 450 GNMETMMIEDKPYAGRELLDVFERFLREAFAHPEETDDLATDMAWYLWTGEYSSLFGKRA 509

Query: 481 MTTFERYYVSDKVTHQERKNPYFKLRDKDDICTALLQEFDL-PKFGHIVNGHTPVKEKNG 539
           MTTFERY++ +K TH+E+KNPY+ LR+ +  C  +L EF L P  GHI+NGHTPVKE  G
Sbjct: 510 MTTFERYFIKEKETHKEKKNPYYYLREDEATCRNILAEFGLNPDHGHIINGHTPVKEIEG 569

Query: 540 EQPIKANGKMLVIDGGFAKGYQKNTGLAGYTLIYNSYGIQLISHLPFTSIEEVLSGTNYI 599
           E PIKANGKM+VIDGGF+K YQ  TG+AGYTL+YNSYG+QL++H  F S  EVLS     +
Sbjct: 570 EDPIKANGKMIVIDGGFSKAYQSTTGIAGYTLLYNSYGMQLVAHKHFNSKAEVLSTGTDV 629

Query: 600 IDTKRLVEEAKDRILVKDTTIGQKLTKEIKDLDHL--YRHFQ 639
           +  KRLV++  +R  VK+T +G++L +E+  L+ L  YR+ +
Sbjct: 630 LTVKRLVDKELERKKVKETNVGEELLQEVAILESLREYRYMK 671
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5486 (GBS197) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 168 (lane 17 & 18; MW 89kDa) and in Figure 169 (lane 2; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 6; MW 99kDa).

Purified Thio-GBS197-His is shown in Figure 244, lane 6.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1766

A DNA sequence (GBSx1873) was identified in *S.agalactiae* <SEQ ID 5487> which encodes the amino acid sequence <SEQ ID 5488>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12719 GB:Z99108 alternate gene name: ygaP~similar to
          hypothetical proteins [Bacillus subtilis]
 Identities = 176/367 (47%), Positives = 240/367 (64%), Gaps = 6/367 (1%)

Query: 3   IKAEIQKLAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIYPERLLE 62
           +K E+ + AK IG+ KIGFTTAD FD L+  L      G  SGFE   IE R+ P+ LL
Sbjct: 55  LKEELIEYAKSIGVDKIGFTTADTFDSLKDRLILQESLGYLSGFEEPDIEKRVTPKLLLP 114

Query: 63  SAKTIISIGVAYPHKLPQQPQKT-SYKRGKITPNSWGLDYHYVVGEKLDRLSKGIEELCR 121
            AK+I++I +AYP ++   P+ T + +RG     SWG DYH V+ EKLD L   ++
Sbjct: 115 KAKSIVAIALAYPSRMKDAPRSTRTERRGIFCRASWGKDYHDVLREKLDLLEDFLKSKHE 174

Query: 122 DFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMFLGELITNLEIEPDKP 181
           D  ++ K+MVDTG L D AVA+RAGIGF  KN ++ + EYGSY++L E+ITN+  EPD P
Sbjct: 175 D--IRTKSMVDTGELSDRAVAERAGIGFSAKNCMITTPEYGSYVYLAEMITNIPFEPDVP 232
```

```
Query: 182 VDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDIC 241
           ++  CG C +CLDACPT  L+  G +NA+RC+SF TQ KG +  EFR KI   +YGCD C
Sbjct: 233 IEDMCGSCTKCLDACPTGALVNPGQLNAQRCISFLTQTKGFLPDEFRTKIGNRLYGCDTC 292

Query: 242 QICCPYNKGINNPLATEI--DPELAQPELIPFLSLSNGQFKEKFGMIAGSWRGKNILQRN 299
           Q  CP NKG +  L  E+  DPE+A+P L P L++SN +FKEKFG ++GSWRGK  +QRN
Sbjct: 293 QTVCPLNKGKDFHLHPEMEPDPEIAKPLLKPLLAISNREFKEKFGHVSGSWRGKKPIQRN 352

Query: 300 AIIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILEFMSNLTLKDE 359
           AI+ALA+  D +A+ +L E++ K+  P+   TA WA+G+I      E LE       KDE
Sbjct: 353 AILALAHFKDASALPELTELMHKDPRPVIRGTAAWAIGKIGDPAYAEELEKALEKE-KDE 411

Query: 360 DSRKELE 366
           +++ E+E
Sbjct: 412 EAKLEIE 418
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5489> which encodes the amino acid sequence <SEQ ID 5490>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3337(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 363/374 (97%), Positives = 367/374 (98%)

Query: 1   MDIKAEIQKLAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIYPERL 60
           M IKAEI+ LAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIY ERL
Sbjct: 18  MTIKAEIKALAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIYTERL 77

Query: 61  LESAKTIISIGVAYPHKLPQQPQKTSYKRGKITPNSWGLDYHYVVGEKLDRLSKGIEELC 120
           LESAKTIISIGVAYPHKLPQQPQKT YKRGKITP+SWGLDYHYVVGEKLDRLSKGIEELC
Sbjct: 78  LESAKTIISIGVAYPHKLPQQPQKTPYKRGKITPSSWGLDYHYVVGEKLDRLSKGIEELC 137

Query: 121 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMFLGELITNLEIEPDK 180
           RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMFLGELITNLEIEPDK
Sbjct: 138 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMFLGELITNLEIEPDK 197

Query: 181 PVDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDI 240
           PVDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDI
Sbjct: 198 PVDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDI 257

Query: 241 CQICCPYNKGINNPLATEIDPELAQPELIPFLSLSNGQFKEKFGMIAGSWRGKNILQRNA 300
           CQICCPYNKGINN  ATEIDPELAQPELIPFLSLSNG+FKEKFGMIAGSWRGKNILQRNA
Sbjct: 258 CQICCPYNKGINNSPATEIDPELAQPELIPFLSLSNGKFKEKFGMIAGSWRGKNILQRNA 317

Query: 301 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILEFMSNLTLKDED 360
           IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEIL FMS+LTLKDED
Sbjct: 318 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILAFMSHLTLKDED 377

Query: 361 SRKELELIRHKWQF 374
           SRKELELIRHKWQF
Sbjct: 378 SRKELELIRHKWQF 391
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1767

A DNA sequence (GBSx1874) was identified in *S.agalactiae* <SEQ ID 5491> which encodes the amino acid sequence <SEQ ID 5492>. This protein is predicted to be peptide chain release factor 2 , fragment (prfB). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4903(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67303 GB:AF017113 putative peptide chain release factor RF-2
          [Bacillus subtilis]
 Identities = 194/336 (57%), Positives = 251/336 (73%), Gaps = 2/336 (0%)

Query: 2   EEEIALLENQMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE- 60
           E  IA L+ QM +P+FWND   AQ    E N LK   +++  + E  +E ++  ++L E
Sbjct: 30  EARIAELDEQMADPEFWNDQQKAQTVINEANGLKDYVNSYKKLNESHEELQMTHDLLKEE 89

Query: 61  -DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLR 119
            D  L+ ELE+ L  L K    +E+ LLLSEPYD NNAILE+HPG+GGTE+QDWG +LLR
Sbjct: 90  PDTDLQLELEKELKSLTKEFNEFELQLLLSEPYDKNNAILELHPGAGGTESQDWGSMLLR 149

Query: 120 MYTRFGNANGFKVEVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSA 179
           MYTR+G   GFKVE LDY  GDEAGIKSVTL  +G NAYG LK+E GVHRLVRISPFDS+
Sbjct: 150 MYTRWGERRGFKVETLDYLPGDEAGIKSVTLLIKGHNAYGYLKAEKGVHRLVRISPFDSS 209

Query: 180 KRRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGI 239
            RRHTSF S EVMPE +D I++++R +DIK+DT+R+ GAGGQ+VN   + VR+TH+PT +
Sbjct: 210 GRRHTSFVSCEVMPEFNDEIDIDIRTEDIKVDTYRASGAGGQHVNTTDSAVRITHLPTNV 269

Query: 240 VVSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKKEITWGSQIRSYVFTP 299
           VV+   +R+Q  NR+RAMKML+AKLYQ   E++  E+D ++G++KEI WGSQIRSYVF P
Sbjct: 270 VVTCQTERSQIKNRERAMKMLKAKLYQRRIEEQQAELDEIRGEQKEIGWGSQIRSYVFHP 329

Query: 300 YTMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRI 335
           Y+MVKDHRTN E+  V  VMDG+I+ FIDAYL+ ++
Sbjct: 330 YSMVKDHRTNTEMGNVQAVMDGDIDTFIDAYLRSKL 365
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5493> which encodes the amino acid sequence <SEQ ID 5494>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 334/337 (99%), Positives = 336/337 (99%)

Query: 1   MEEEIALLENQMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE 60
           +EEEIALLEN MTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE
Sbjct: 1   LEEEIALLENHMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE 60

Query: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLRM 120
           DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLRM
Sbjct: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLRM 120
```

```
Query: 121 YTRFGNANGFKVEVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSAK 180
           YTRFGNANGFK+EVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSAK
Sbjct: 121 YTRFGNANGFKIEVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSAK 180

Query: 181 RRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGIV 240
           RRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGIV
Sbjct: 181 RRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGIV 240

Query: 241 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKKEITWGSQIRSYVFTPY 300
           VSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKKEITWGSQIRSYVFTPY
Sbjct: 241 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKKEITWGSQIRSYVFTPY 300

Query: 301 TMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRIED 337
           TMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRIED
Sbjct: 301 TMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRIED 337
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1768

A DNA sequence (GBSx1875) was identified in *S.agalactiae* <SEQ ID 5495> which encodes the amino acid sequence <SEQ ID 5496>. This protein is predicted to be cell-division ATP-binding protein (ftsE). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3928(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67262 GB:AF017113 cell division ATP-binding protein [Bacillus subtilis]
 Identities = 138/228 (60%), Positives = 179/228 (77%)

Query: 3   LIEMSGVTKKYRRSTTALRNLNLSIQQGEFVYLVGPSGAGKSSLIRLLYREEKLSSGRLK 62
            +IEM  V K Y     AL  ++++I  GEFVY+VGPSGAGKS+ I+++YREEK + G++
Sbjct: 1   MIEMKEVYKAYPNGVKALNGISVTIHPGEFVYVVGPSGAGKSTFIKMIYREEKPTKGQIL 60

Query: 63  VGEFNLNKLKRRQIPILRRSIGVVFQDYKLLPTKTVYENVAFAMQVIGAKRRHIKKRVPE 122
            +   +L  +K ++IP +RR IGVVFQD+KLLP  TV+ENVAFA++VIG +   IKKRV E
Sbjct: 61  INHKDLATIKEKEIPFVRRKIGVVFQDFKLLPKLTVFENVAFALEVIGEQPSVIKKRVLE 120

Query: 123 VLELVGLKHKMRSFPTQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEIAWEIMHLL 182
           VL+LV LKHK R FP QLSGGEQQRV+IAR+IVNNP ++IADEPTGNLDP+ +WE+M  L
Sbjct: 121 VLDLVQLKHKARQFPDQLSGGEQQRVSIARSIVNNPDVVIADEPTGNLDPDTSWEVMKTL 180

Query: 183 ERINLQGTTVLMATHNSQIVNTLRHRVIEIEAGSVIRDEEKGEYGYHD 230
           E IN +GTTV+MATHN +IVNT++ RVI IE G ++RDE +GEYG +D
Sbjct: 181 EEINNRGTTVVMATHNKEIVNTMKKRVIAIEDGIIVRDESRGEYGSYD 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5497> which encodes the amino acid sequence <SEQ ID 5498>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3728(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 191/230 (83%), Positives = 214/230 (93%)

Query: 1   MALIEMSGVTKKYRRSTTALRNLNLSIQQGEFVYLVGPSGAGKSSLIRLLYREEKLSSGR 60
            MALIEMSGVTKKYRRSTTALR++N+S+ QGEFVYLVGPSGAGKS+ I+LLYREE+L++G+
Sbjct: 1   MALIEMSGVTKKYRRSTTALRDVNVSVNQGEFVYLVGPSGAGKSTFIKLLYREEQLTTGK 60

Query: 61  LKVGEFNLNKLKRRQIPILRRSIGVVFQDYKLLPTKTVYENVAFAMQVIGAKRRHIKKRV 120
            L VGEFNL KLK R +PILRR IGVVFQDYKLLP KTV+ENVA+AM+VIG KRRHIKKRV
Sbjct: 61  LYVGEFNLTKLKARDVPILRRHIGVVFQDYKLLPRKTVFENVAYAMEVIGEKRRHIKKRV 120

Query: 121 PEVLELVGLKHKMRSFPTQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEIAWEIMH 180
            PEVL+LVGLKHKMRSFP+QLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEI+WEIM
Sbjct: 121 PEVLDLVGLKHKMRSFPSQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEISWEIMQ 180

Query: 181 LLERINLQGTTVLMATHNSQIVNTLRHRVIEIEAGSVIRDEEKGEYGYHD 230
            LLERIN+QGTT+LMATHNS IVNT RHRV+ IE G ++RDEEKG+YGY D
Sbjct: 181 LLERINVQGTTILMATHNSHIVNTFRHRVVAIEDGRIVRDEEKGDYGYDD 230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1769

A DNA sequence (GBSx1876) was identified in *S.agalactiae* <SEQ ID 5499> which encodes the amino acid sequence <SEQ ID 5500>. This protein is predicted to be ftsE protein (ftsX). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77    Transmembrane  296 - 312 ( 291 - 322)
   INTEGRAL    Likelihood = -9.24    Transmembrane  203 - 219 ( 198 - 228)
   INTEGRAL    Likelihood = -6.16    Transmembrane   49 -  65 (  40 -  68)
   INTEGRAL    Likelihood = -3.40    Transmembrane  255 - 271 ( 252 - 273)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9629> which encodes amino acid sequence <SEQ ID 9630> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67264 GB:AF017113 cell division protein [Bacillus subtilis]
 Identities = 112/311 (36%), Positives = 182/311 (58%), Gaps = 31/311 (9%)

Query: 27  RHFWESLKNLKRNFWMTFASVTSVTITLLLVGLFSSVLLNVEKLTTDVSGNFTISAFLNV 86
            RH  ES K+L RN WMTFAS+++VT+TL+LVG+F  ++LN+  + T+      I   +++
Sbjct: 7   RHLRESFKSLGRNTWMTFASISAVTVTLILVGVFLVIMLNLNNMATNAEKQVEIKVLIDL 66

Query: 87  DSTDAQKQVKDKDGKLKDNPDYHKVYDKIKRISGVEKVTYSSKAEQLKEVQKEYGSDVID 146
             +          D K +D     K+ + IK + G++ VT+SSK ++L ++   +G
Sbjct: 67  TA----------DQKAQD-----KLQNDIKELKGIQSVTFSSKEKELDQLVDSFGDSGKS 111

Query: 147 DTYKDA---LLDVYVVGTSSAKVSKSVSEAIGRIEGV---DYTKEPIDST-KLSNLTDNI 199
             T KD    L D +VV T+    + +V++ I +++ V    Y KE +    K+  ++ NI
Sbjct: 112 LTMKDQENPLNDAFVVKTTDPHDTPNVAKKIEKMDHVYKVTYGKEEVSRLFKVVGVSRNI 171

Query: 200 RIWGFGGVALLIVL---AIFLISNTIRMSIMSRRTDIEIMRLVGAKNSYIRGPFFFEGAW 256
```

```
                 G+AL+I L   A+FLISNTI+++I +RR +IEIM+LVGA N +IR PFF EG
Sbjct: 172 ------GIALIIGLVFTAMFLISNTIKITIFARRKEIEIMKLVGATNWFIRWPFFLEGLL 225

Query: 257 VGILGAIVPSLIFYFGYQFVFNKFNPKFETSHVSLYPMDIMVPAIIGGMVIIGIIIGSLG 316
           +G+ G+++P  +    YQ+V     PK + S VSL P +  V  +   ++ IG +IG  G
Sbjct: 226 LGVFGSVIPIALVLSTYQYVIGWVVPKVQGSFVSLLPYNPFVFQVSLVLIAIGAVIGVWG 285

Query: 317 SVLSMRRYLKI 327
           S+ S+R++L++
Sbjct: 286 SLTSIRKFLRV 296
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5501> which encodes the amino acid sequence <SEQ ID 5502>. Analysis of this protein sequence reveals the following:

```
        Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.70   Transmembrane  195 - 211 ( 189 - 219)
   INTEGRAL    Likelihood = -6.74   Transmembrane   39 -  55 (  30 -  58)
   INTEGRAL    Likelihood = -5.52   Transmembrane  294 - 310 ( 288 - 314)
   INTEGRAL    Likelihood = -1.49   Transmembrane  246 - 262 ( 245 - 263)

----- Final Results -----
               bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC67264 GB:AF017113 cell division protein [Bacillus subtilis]
 Identities = 117/311 (37%), Positives = 184/311 (58%), Gaps = 19/311 (6%)

Query: 11  MIRYFFRHIWESIKNLKRNFWMTFASVSMVAVTLTLVGVFAATLLNIQRVASGVENNVHI 70
           MI+   RH+ ES K+L RN WMTFAS+S V VTL LVGVF   +LN+  +A+  E  V I
Sbjct: 1   MIKILGRHLRESFKSLGRNTWMTFASISAVTVTLILVGVFLVIMLNLNNMATNAEKQVEI 60

Query: 71  NTYLQVDSTDAAKVIQNTAGEPVNNDNYHSVYDKIAQIKGVKKITFSSKDEQLKKLQETL 130
              + + +   A+       + + ND        I ++KG++ +TFSSK+++L +L ++
Sbjct: 61  KVLIDLTADQKAQ-------DKLQND--------IKELKGIQSVTFSSKEKELDQLVDSF 105

Query: 131 GDVWN---MYDQDTNPLQDIYLIETQTPKQVKAITKKIRTIEGVEAADYGGINSDKLFKF 187
           GD      M DQ+ NPL D ++++T  P     + KKI  ++ V    YG     +LFK
Sbjct: 106 GDSGKSLTMKDQE-NPLNDAFVVKTTDPHDTPNVAKKIEKMDHVYKVTYGKEEVSRLFKV 164

Query: 188 STLIQTWGLIGTAMLLFVAVFLISNTIRMTIMSRKRDIEIMRLVGAKNSYIRGPFFFEGA 247
             + +  G+     L+F A+FLISNTI++TI +R+++IEIM+LVGA N +IR PFF EG
Sbjct: 165 VGVSRNIGIALIIGLVFTAMFLISNTIKITIFARRKEIEIMKLVGATNWFIRWPFFLEGL 224

Query: 248 WVGLLGAVLPSLLIYYGYDLVYKHFAQELQRNNLSMYPLDPYVYYLIGALFVIGIMIGSL 307
            +G+ G+V+P  L+   Y  V      ++Q + +S+ P +P+V+ +   L  IG +IG
Sbjct: 225 LLGVFGSVIPIALVLSTYQYVIGWVVPKVQGSFVSLLPYNPFVFQVSLVLIAIGAVIGVW 284

Query: 308 GSVLSMRRYLK 318
           GS+ S+R++L+
Sbjct: 285 GSLTSIRKFLR 295
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 173/318 (54%), Positives = 238/318 (74%), Gaps = 5/318 (1%)

Query: 13  MKRRENMVIMIN-FFRHFWESLKNLKRNFWMTFASVTSVTITLLLVGLFSSVLLNVEKLT 71
           MK++E MV MI  FFRH WES+KNLKRNFWMTFASV+ V +TL LVG+F++ LLN++++
Sbjct: 2   MKKKEIMVTMIRYFFRHIWESIKNLKRNFWMTFASVSMVAVTLTLVGVFAATLLNIQRVA 61

Query: 72  TDVSGNFTISAFLNVDSTDAQKQVKDKDGKLKDNPDYHKVYDKIKRISGVEKVTYSSKAE 131
           + V  N  I+ +L VDSTDA K +++  G+  +N +YH VYDKI +I GV+K+T+SSK E
Sbjct: 62  SGVENNVHINTYLQVDSTDAAKVIQNTAGEPVNNDNYHSVYDKIAQIKGVKKITFSSKDE 121
```

```
Query: 132 QLKEVQKEYGSDVID--DTYKDALLDVYVVGTSSAKVSKSVSEAIGRIEGVDYTKEP-ID 188
           QLK++Q+  G DV +  D   + L D+Y++ T + K  K++++ I  IEGV+      I+
Sbjct: 122 QLKKLQETLG-DVWNMYDQDTNPLQDIYLIETQTPKQVKAITKKIRTIEGVEAADYGGIN 180

Query: 189 STKLSNLTDNIRIWGFGGVALLIVLAIFLISNTIRMSIMSRRTDIEIMRLVGAKNSYIRG 248
           S KL   +  I+ WG  G A+L+ +A+FLISNTIRM+IMSR+ DIEIMRLVGAKNSYIRG
Sbjct: 181 SDKLFKFSTLIQTWGLIGTAMLLFVAVFLISNTIRMTIMSRKRDIEIMRLVGAKNSYIRG 240

Query: 249 PFFFEGAWVGILGAIVPSLIFYFGYQFVFNKFNPKFETSHVSLYPMDIMVPAIIGGMVII 308
           PFFFEGAWVG+LGA++PSL+ Y+GY  V+  F  + + +++S+YP+D  V  +IG + +I
Sbjct: 241 PFFFEGAWVGLLGAVLPSLLIYYGYDLVYKHFAQELQRNNLSMYPLDPYVVYLIGALFVI 300

Query: 309 GIIIGSLGSVLSMRRYLK 326
           GI+IGSLGSVLSMRRYLK
Sbjct: 301 GIMIGSLGSVLSMRRYLK 318
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1770

A DNA sequence (GBSx1877) was identified in *S.agalactiae* <SEQ ID 5503> which encodes the amino acid sequence <SEQ ID 5504>. This protein is predicted to be carboxymethylenebutenolidase-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF10898 GB:AE001979 carboxymethylenebutenolidase-related
           protein [Deinococcus radiodurans]
 Identities = 65/183 (35%), Positives = 98/183 (53%), Gaps = 3/183 (1%)

Query: 56  SKGKVKANIIFYQGALVEEEAYSQLARDLADKGDNTYILKTPLNLPVLSPHKAKTIINQN 115
           +  +VK  ++FY G  V  +AY  L R LA +G  T I   PL+L +    +A+ +I +
Sbjct: 100 ASAEVKTLLVFYPGGRVRPQAYEWLGRALAVRGVQTVIPAFPLDLAITGTERAEGLIARY 159

Query: 116 HL-TNVYLAGHSLGGVVASQNAKVAP--VRGLILLASYPSRKSDLSHKNLRVLSITASND 172
                V LAGHSLGG VA+Q A + P  + GL+LLA+YP+   +L       LS+ A  D
Sbjct: 160 GAGKRVVLAGHSLGGTVAAQYAALRPDKIDGLLLLAAYPAPNVNLHDARFPALSLLAEKD 219

Query: 173 HILNWEKYEEAKKRLPNSSTFRTIVGGNHSRFGNYGHQKGDGKATLSHKSSEKQLATFIS 232
            + +        +RLP ++    + G  HS FG YG Q+GDG  T+S   +E+++   +
Sbjct: 220 GVADAGLVRGGLERLPKNTRLTVLPGAVHSFFGRYGPQQGDGVPTVSRARAEREIVQAVE 279

Query: 233 NFI 235
            FI
Sbjct: 280 TFI 282
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5504 (GBS158) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 5; MW 52kDa).

The GBS158-GST fusion product was purified (Figure 113; see also Figure 201, lane 4) and used to immunise mice (lane 1+2 product; 14.5µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1771

A DNA sequence (GBSx1878) was identified in *S.agalactiae* <SEQ ID 5505> which encodes the amino acid sequence <SEQ ID 5506>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0281(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06539 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 83/197 (42%), Positives = 114/197 (57%), Gaps = 4/197 (2%)

Query: 35  NTYYLVNDQAV-ILIDPGSNGQEIIAKIKSFEKPLVAILLTHTHYDHIFSLDLVRDTFDN 93
            N Y   NDQ   I+ DPG   +++I  ++  +   +AILLTH H+DHI +++ VR+TF +
Sbjct: 14  NWYIQTNDQGEGIIFDPGGEVEKLITWLRDRQITPLAILLTHAHFDHIGAVEDVRNTF-H 72

Query: 94  PPVYVSEKEAAWLSSPDDNLSGLGRHDDIINVIARPAENFFKLKQPYQLNGFEFTVLPTP 153
             PVY+ E E  WL  P  N S L      I   AR AE+    +Q   +  F + VL TP
Sbjct: 73  IPVYIHENEKEWLIDPQRNGSSLFIPGSSIK--AREAEHLITGEQDLSIGSFSYQVLETP 130

Query: 154 GHSWGGVSFVFHSDELVVTGDALFRETIGRTDLPTSNFEDLITGIRQELFTLPSHYSVHP 213
            GHS G +S+    D++V +GDALF  +IGRTDLP  + + L+  I  +L  LP   +V
Sbjct: 131 GHSPGSLSYYAKEDKIVFSGDALFAGSIGRTDLPGGDHQLLLDSIHDKLLELPEDTTVAS 190

Query: 214 GHGMNTTIGHEKNFNPF 230
            GHG  TTIGHE + NPF
Sbjct: 191 GHGPTTTIGHEMDGNPF 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5507> which encodes the amino acid sequence <SEQ ID 5508>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0407(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 217/231 (93%), Positives = 224/231 (96%)

Query: 1   MPFIFRHSFFNKVLIFWYTIIMKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK 60
            +PFIFR+SFFNKVLIFWYTI+MKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK
Sbjct: 1   LPFIFRYSFFNKVLIFWYTILMKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK 60

Query: 61  IKSFEKPLVAILLTHTHYDHIFSLDLVRDTFDNPPVYVSEKEAAWLSSPDDNLSGLGRHD 120
            IKSFEKPLVAILLTHTHYDHIFSLDLVRD FD+PPVYVSEKEAAWLSSPDDNLSGLGRHD
```

```
Sbjct: 61  IKSFEKPLVAILLTHTHYDHIFSLDLVRDAFDHPPVYVSEKEAAWLSSPDDNLSGLGRHD 120

Query: 121 DIINVIARPAENFFKLKQPYQLNGFEFTVLPTPGHSWGGVSFVFHSDELVVTGDALFRET 180
           DII VIARPAENFFKLKQPYQLNGFEFTVLPT GHSWGGVSFVFHSDELVVTGDALFRET
Sbjct: 121 DIITVIARPAENFFKLKQPYQLNGFEFTVLPTSGHSWGGVSFVFHSDELVVTGDALFRET 180

Query: 181 IGRTDLPTSNFEDLITGIRQELFTLPSHYSVHPGHGMNTTIGHEKNFNPFF 231
           IGRTDLPTSNFEDLITGIRQELFTLP+HY V+PGHG +TTI HEKN NPFF
Sbjct: 181 IGRTDLPTSNFEDLITGIRQELFTLPNHYRVYPGHGPSTTICHEKNANPFF 231
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1772

A DNA sequence (GBSx1879) was identified in *S.agalactiae* <SEQ ID 5509> which encodes the amino acid sequence <SEQ ID 5510>. This protein is predicted to be acetoin reductase (fabG). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1596(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9631> which encodes amino acid sequence <SEQ ID 9632> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC48769 GB:U71200 acetoin reductase [Bos taurus]
 Identities = 162/254 (63%), Positives = 188/254 (73%), Gaps = 2/254 (0%)

Query: 12  KVAIVTGAGQGIGFAIAKRLHADGFKIGVLDYNEETAQAAVDKLSPED--AVAVVADVSK 69
           KVA+VTG  QGIG AI   L ADGF + V D NE  ++     +      A+AV  DVS
Sbjct: 4   KVAMVTGGAQGIGEAIVXXLSADGFAVAVADLNEAKSKXVATDIEKNGGTAIAVKLDVSD 63

Query: 70  RDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITEEQFEKAFAINVGGTIWGSQAAQ 129
           R+  F A ++V +  G  +V+VNNAG+ PTTP+DTIT E F+K + INV G IWG QAA
Sbjct: 64  REGFFAAVKEVAEKLGGFDVLVNNAGLGPTTPIDTITPELFDKVYHINVAGDIWGIQAAV 123

Query: 130 KHFRELGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASEGITVNAYAP 189
           + F++ G+GGKIINATSQAG  GNPNL++Y  TKFAVR +T   A+DLA + ITVNAYAP
Sbjct: 124 EQFKKNGNGGKIINATSQAGVVGNPNLSLYSSTKFAVRCLTPVAARDLAEQNITVNAYAP 183

Query: 190 GIVKTPMMFDIAHEVGKNAGKDDEWGMEQFAKDITLKRLSEPEDVANAVGFLAGDDSNYI 249
           GIVKTP  FDIAHEVGKNAGKDDEWGM+ FAKDI LKRLSEPEDVA AV FLAG DSNYI
Sbjct: 184 GIVKTPXXFDIAHEVGKNAGKDDEWGMQTFAKDIALKRLSEPEDVAAAVAFLAGPDSNYI 243

Query: 250 TGQTIVVDGGMVFH 263
           TGQTI VDGGM FH
Sbjct: 244 TGQTIEVDGGMQFH 257
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5511> which encodes the amino acid sequence <SEQ ID 5512>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1131(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```